Research Policy 43 (2014) 333–348

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

International research networks in pharmaceuticals: Structure and dynamics

Uwe Cantner[a,b], Bastian Rake[c,*]

[a] Friedrich Schiller University Jena, Faculty of Economics and Business Administration, Carl-Zeiss-Strasse 3, D-07743 Jena, Germany
[b] University of Southern Denmark, Department of Marketing and Management, Campusveij 55, DK-5230 Odense M, Denmark
[c] University of Passau, Chair of Organization, Technology Management and Entrepreneurship (OTE), Innstrasse 27, D-94032 Passau, Germany

ARTICLE INFO

Article history:
Received 16 November 2011
Received in revised form 22 October 2013
Accepted 28 October 2013
Available online 4 December 2013

JEL classification:
R10
O31

Keywords:
International cooperation
Pharmaceuticals
Research networks
Network dynamics
MRQAP

ABSTRACT

Knowledge production and scientific research have become increasingly more collaborative and international, particularly in pharmaceuticals. We analyze this tendency in general and tie formation in international research networks on the country level in particular. Based on a unique dataset of scientific publications related to pharmaceutical research and applying social network analysis, we find that both the number of countries and their connectivity increase in almost all disease group specific networks. The cores of the networks consist of high income OECD countries and remain rather stable over time. Using network regression techniques to analyze the network dynamics our results indicate that accumulative advantages based on connectedness and multi-connectivity are positively related to changes in the countries' collaboration intensity whereas various indicators on similarity between countries do not allow for unambiguous conclusions.

1. Introduction

Collaboration between different actors has become an increasingly more important mode of knowledge generation in almost all scientific disciplines (Wuchty et al., 2007). Particularly in science-based fields and research areas with rapidly developing and widely distributed knowledge bases, no single actor has the ability to keep pace with the scientific and technological advances in all areas. Consequently, increasing collaboration within 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). In this paper we are interested in the dynamics of collaboration networks in general and tie formation herein in particular.

We pursue our analysis for pharmaceuticals, where innovation is based on scientific advances and thus clearly connected to basic and applied research (Lim, 2004). Pharmaceutical innovation can be seen the result 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). Since this field is characterized by a complex, expanding and dispersed knowledge base, 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). Its structure and the actors' positions therein determine the actors' access to relevant sources of knowledge and may therefore have consequences for their innovative activities and performance (Powell et al., 1999; Walker et al., 1997; Kogut et al., 1994).

Against this background of a specific research area, pharmaceuticals, and following the analytical design of Wagner and Leydesdorff (2005a), we pursue an analysis on the country level, implying that the collaborating actors within the network are countries. Our deviation from an analysis on the level of individual actors is justified by the presumption that for the dynamics of international collaboration networks in pharmaceutical research the country level embraces an additional and important influence on the formation of and the changes in those networks. Looking at the research and collaboration activities in pharmaceuticals worldwide as measured by publications and co-authored publications we, first, find the OECD countries to be the main actors in both categories. Secondly, these countries show a continuous reinforcement of the co-authorship ties among each other over time combined with a slightly but steadily increasing widening out of collaborative

* Corresponding author. Tel.: +49 851 509 2483.
E-mail addresses: uwe.cantner@uni-jena.de (U. Cantner), bastian.rake@uni-passau.de (B. Rake).

0048-7333/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.respol.2013.10.016

activities to other countries. For OECD countries this, third, is evidenced by a growth of the number of internationally collaborated publications exceeding the growth of the number of non-collaborated ones: their growth rates of the number of internationally collaborated publications are about 37% which is almost twice as high as their growth rates of non-collaborated publications. Fourth, for non-OECD countries, starting from a much lower number of publications, the growth rates of publications are much higher, between 60% and 95%, with only a minor difference between the growth of the number internationally collaborated and non-collaborated publications for this group.[1] This, fifth, leads to a share of collaborative in all publications which increases for OECD countries from almost 17% to 22% and which remains nearly stable for non-OECD countries between 20% and 21%. Taken these five observations for worldwide research in pharmaceuticals together, OECD countries are dominant in these activities, show a higher inclination for, and put more emphasis on international scientific collaboration. Hence, from a research network perspective, OECD countries tend to be at the core of those networks and should be considered the main drivers of network dynamics. We consider these patterns as an indication for country level factors having a prominent effect on the structure and dynamics of the international pharmaceutical research network.

Among the country level factors, in our context, the national innovative capacity matters most. It is considered as the binding frame for research as well as innovation activities and their very structures influence the long-term ability of a country to generate and commercialize innovative technologies (Furman et al., 2002). Countries differ in their national innovative capacity. Hence, the exchange of knowledge and approaches for problem solving among organizations embedded in national contexts with country-specific scientific and technological advantages can be seen as a driver for international collaborations (Bartholomew, 1997; Shan and Hamilton, 1991; Dosi et al., 1990). Although the national innovation systems are connected among each other, the development of the pharmaceutical industry in general and the development of the related research networks in particular are closely connected to the structure of national institutions (Henderson et al., 1999). Differences in the national institutional setting and the national innovative capacity may at least partly explain the considerable differences of research activities in different disease groups on the country level (cf. Furman et al., 2006).

International scientific collaboration between countries can be seen as a self-organizing network as suggested in Wagner and Leydesdorff (2005a). The position of each country within the network is determined by economic, social, political, and cultural factors. These factors include the priorities of scientists and policy makers to conduct research in particular fields and to collaborate with different partners (Miquel and Okubo, 1994). They can be considered as part of the national innovative capacity of a country that influences by itself scientific collaboration in lower order subsystems, such as pharmaceuticals, and induces the dynamics of the cross-country network. These subsystems contribute to the dynamics on the international level while they are at the same time affected by the dynamics on the international level (Wagner and Leydesdorff, 2005a). Within this analytical context, we explore the structure and the dynamics of international research collaboration networks for different disease groups in pharmaceutical research.

We use social network analysis to investigate 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. Therefore, we test whether the connectedness of countries, the similarity of countries or their degree of multi-connectivity are the driving factors of tie formation within the networks. In order to investigate the network dynamics, we draw upon multiple regression analysis for dyadic data (Butts and Carley, 2001; Krackhardt, 1988). More precisely, we apply the multiple regression quadratic assignment procedure (MRQAP) with double semi-partialing (DSP) as proposed by Dekker et al. (2007), which is particularly robust against multicollinearity and network-autocorrelation.

Our empirical analysis is performed on a unique dataset of publications in scientific journals related to pharmaceutical research. We analyze three periods, 1998–2000, 2002–2004, and 2006–2008. A first 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 countries in the core, only a few non-OECD countries have managed to become part of the center of the international research community. Our descriptive network statistics indicate increasing cross-county collaboration in almost all disease groups.

Our regression results reveal that dyadic tie formation and break-up is positively related to the amount of previous collaboration. This finding may indicate an accumulative advantage based on the degree of connectedness of a country in a network. We do not find a clear-cut association between differences in the visibility of two countries, as another proxy for connectedness, and the change in the number of research collaborations. Moreover, similarity of two countries in terms of income groups and language has no unambiguous association with the changes in the number of collaborations. Country differences in the strength of their research sectors are negatively related to the change of their bilateral collaborations. Multi-connectivity, in terms of other countries connecting two countries, is positively related to changes in the number of collaborations between them, whereas we find a negative association for the number of shortest paths between two countries. From a policy perspective, our results suggest that supporting international research collaborations and investments strengthening countries' scientific systems may help countries located at peripheral positions in the network to get access to the relevant sources of knowledge and to overcome liabilities of unconnectedness.

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 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. Internationalization of research networks

Research and innovation activities worldwide appear to be performed more and more in collaboration. With respect to scientific collaboration, there is a large body of evidence for an increasing amount 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. This growth in scientific collaboration activities is not bound within national boundaries but shows an international outreach. As Adams et al. (2005) show, on a large sample of publications originating in U.S research universities, the rate of

[1] This trend is prevalent for the comparison of all three sub-periods analyzed in this study, 1998–2000, 2002–2004, and 2006–2008. We obtain qualitatively similar results if we look at the biggest contributors in the number of publications in the sample instead of OECD countries.

national collaborations more than doubled and the rate of international collaborations increased five-fold from the 1980s to the late 1990s. These results are in line with 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). Increasing collaboration has not only been observed in science, but also with respect to R&D and innovative activities in general. Hagedoorn (2002) finds an increasing number of R&D alliances since the 1980s that can be found in a diverse set of industries, such as the computer, semiconductor, chemical and footwear industries (e.g. Boschma and ter Wal, 2007; Ahuja, 2000; Saxenian, 1991). Moreover, also collaborative R&D activities show an increasing level of internationalization (Guellec and van Pottelsberghe de la Potterie, 2001; Granstrand, 1999).

A prominent case in terms of collaboration in science and innovation is the pharmaceutical industry. In this industry the R&D process is based on a diverse set of knowledge from various scientific disciplines. The rapid growth of an increasingly codified and abstract knowledge base (Gambardella, 1995; Arora and Gambardella, 1994) and its dispersion among a broad variety of actors induced a pronounced trend towards collaboration and network formation (Powell et al., 2005; McKelvey et al., 2004). Organizations (including established pharmaceutical companies, biotechnology firms, universities, public research institutes, and venture capitalists) with complementary resources and competencies are involved in the networks of collaborative relations and innovations are the ultimate outcome of their interaction and collaboration (e.g. Pisano, 1991; Orsenigo, 1989). Numerous studies have described and visualized the growth of R&D partnerships between these different types of organizations (e.g. Roijakkers and Hagedoorn, 2006; Powell et al., 2005). The complementarity of assets and key competencies of biotechnology firms and pharmaceutical companies has been found to be an important driver of the growth of R&D partnerships on the organizational level (e.g. Senker and Sharp, 1997). The international dimension of collaboration in the pharmaceutical industry is particularly pronounced when biotechnological knowledge is involved and regionally clustered organizations 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 on 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).

The observed increasing internationalization of collaborative activities in scientific research and innovation has certainly its origin in the collaboration decisions of the cooperating organizations. Above and beyond that, however, there appears to be a rather aggregated dimension involved, the level of countries. Looking at co-publication activities in science in general reveals an expansion in the number of involved countries and the connections among them (Wagner and Leydesdorff, 2005a). However, not all countries are connected to the core, 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). With respect to R&D collaborations in general, the majority of alliances worldwide are geographically concentrated among North America, Europe, Japan, and South Korea (Hagedoorn, 2002).

For the pharmaceutical industry a similar pattern of internationalization emerges which seems to be even more pronounced, particularly with respect to scientific collaboration (Calero et al., 2007). U.S. based companies, universities, and public research organizations have been dominating the industry since the emergence of biotechnology in the 1970s. Hence, linkages to the U.S. have been important means for European organizations to increase their capabilities and competitiveness (Sharp and Senker, 1999). An analysis 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).

Based on this literature and evidence, we are interested in the international dimension of research networks and we analyze them in terms of connections between countries. Quite generally, a network in an economic sense is composed of heterogeneous actors, in our case countries, the relationships among them and other contextual features that affect actors' behavior and decisions, as well as the generation and application of knowledge. Concerning actors involved quite generally, they have different knowledge and competencies, different rules of action, and different incentives and motivations. They are linked to one another 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).

In this paper, we follow Wagner and Leydesdorff (2005a) and refer to countries as actors in the networks we analyze. The edges between the countries represent co-authorship relationships between authors located different countries. The "behavior" of a country within a network reflects an "aggregate behavior" which is determined by economic, social, political and cultural factors. These factors include the priorities of scientists, organizations, and policy makers to conduct research in particular fields and to collaborate with different international partners (Miquel and Okubo, 1994). In this sense the "aggregate behavior" we refer to is considered as being quite closely related to the national innovative capacity of a country. Certainly, we cannot observe an "aggregate behavior" directly, but we consider a country's pattern of co-authorship relations being the outcome of an "aggregate behavior" and we track these patterns over time.

2.2. Network dynamics

Based on the increasing importance of international scientific collaboration, we analyze changes in collaboration networks over time. For that we anchor our analysis onto a theory that draws on network theory in general and on a theory on the dynamics of networks in particular. Network theory (e.g. Burt, 1992; Granovetter, 1973) suggests a relational approach to understand why certain units are connected. It considers the properties of their relations and hence a comparison (similarities and dissimilarities) between them as relevant, instead of the properties of the units itself. Hence, we refer to similarities and dissimilarities of network units in terms of the properties of ties between them, such as the number of ties, as well as in terms of relational states and events that are not directly dyadic, such as economic similarities (Borgatti and Halgin, 2011).

These elements of network theory are taken over to our approach on the dynamics of networks (e.g. Ahuja et al., 2012; Rivera et al., 2010; Powell et al., 2005), and to the related changes herein. Referring to a notion of change as suggested in evolutionary economics, we consider processes that lead to a transformation of a system from within (Witt, 2008; Schumpeter, 1912). More formally, changes to be observed in t are not independent from past events or states in $t-1$. In the context of collaboration networks, our evolutionary view implies that the actors' positions, their connections and their similarities within the network in t influence the ongoing formation and breaking-up of ties (Kenis and Knoke, 2002).

The decisions of the formation or breaking up of new ties are driven by actors' ambitions to maintain or reconfigure their position within the network in order to benefit from opportunities created by the network structure (Zaheer and Soda, 2009). In a scientific co-author network these benefits arise from using, integrating and recombining the knowledge and expertise of another actor; the attractiveness of this potential partner depends positively on his knowledge and expertise as well as on how easy he can be accessed and understood. Respective kinds and degrees of this attractiveness will be introduced below.

We apply our theory of network formation and development to the case of inter-country linkages in pharmaceutical co-publications. Obviously, individual researchers decide on these collaborations. Aggregating these researchers and their collaboration decisions on the country level is based on the assumption that scientists of the same country behave and decide in a very much common way, following Wagner and Leydesdorff (2005a). These commonalities can for example be related to country specific styles and forms of organizing research and science. Our main question in the analysis is on how the network structure in previous periods affects the interaction structure among countries in subsequent periods. We focus on several determinants of this network dynamics which is mainly driven by changes in the number of dyadic relationships. In this paper, we concentrate on the dyadic concept of connectedness of network actors, on relational states indicating similarity of network actors, and multi-connectivity of network actors. We apply these concepts to analyze and understand the development of cross-country co-author networks in pharmaceutical research.

2.2.1. Connectedness

Our first determinant or mechanism driving new tie formation in t is the connectedness of network actors indicating their attractiveness as potential cooperation partners and given by the total number of dyadic relationships a network actor holds in $t-1$. This number indicates a degree of attractiveness in terms of knowledge and expertise others attached to this actor. In relational terms connectedness shows importance for endogenous network formation (Powell et al., 2005) for which we distinguish two ways. A first one is related to a network's core-periphery structure with highly connected core actors on the one hand and low connected peripheral actors as well as (in t) not connected new actors on the other. In those structures dynamics of degree (Rivera et al., 2010) may work and in accordance with the Barabási-Albert model (Barabási and Albert, 1999). One can expect tie formation in terms of preferential attachment based on relative connectedness: New and less well-connected actors establish ties preferably to well-connected incumbents. Put differently, highly connected core actors in $t-1$ are more likely to attract new ties in t. The second way connectedness is related to tie formation in t looks at the number of dyads two network actors hold in common in $t-1$ – implying a kind of repetition dynamics (Rivera et al., 2010). This number of common dyads evidences the attractiveness of the collaboration between the two network actors. Further tie formation between the two appears to be positively related to this type of bilateral connectedness.

Both mechanism, via relative connectedness and via bilateral connectedness, tracked over time are responsible for a "rich-get-richer" phenomenon in which network incumbents (early entrants) increase their connectivity at the (relative) expense of newcomers. A small number of actors shows a high number of ties within the network, whereas the vast majority of actors has relatively few ties. As a consequence, 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).

Empirical analyses on various levels support the conjecture of connectedness driving new tie formation. The dynamics in core-periphery structures has been observed for scientific co-authorship relationships in different disciplines (Wagner and Leydesdorff, 2005b; Jeong et al., 2003; Newman, 2001). Focusing on the firm-level, Orsenigo et al. (1998) show that the network of collaborative R&D agreements in the pharmaceutical industry (after the emergence of biotechnology) expands by keeping structural properties rather stable. Especially the non-deformation of the core-periphery structure and a low propensity to collaborate among firms of similar age indicate that the number of ties an actor has established may have been the driving force in the evolution of the network (Ter Wal and Boschma, 2009).

Based on these arguments and evidence we hypothesize that connectedness is related to changes in the dyadic relationships (tie formation and break up) also in the cross-country co-author network of pharmaceutical research. First we explore whether less well connected countries establish preferably linkages to those countries that are already well connected within the network. In this case, the difference in the number of collaborative linkages to other countries should be related to changes in the number of collaborations between two countries. We secondly investigate whether the number of previous dyadic co-author relationships between two countries is positively related to changes in the ties between them. We summarize these conjectures in Hypotheses 1a and 1b.

Hypothesis 1a. The difference in degree of connectedness (relative connectedness) between two countries is positively related to changes in the number of cross-country co-author collaborations between them.

Hypothesis 1b. The number of common dyads between two countries (bilateral connectedness) is positively related to changes in the number of cross-country co-author collaborations between them.

2.2.2. Similarity

In most real world networks, the tendency to connect to highly connected actors is not as high as theoretical models predict. Two explanations are given. A first one argues that the number of connections an actor can meaningfully maintain is limited. For our cross-country network based on co-authorships this argument is difficult to apply or defend. However, a second explanation forwards that dimensions other than sheer dyadic connectedness have to be considered. In this sense, actors are attracted by those with the highest degree of connectedness, but prefer to connect to proximate or similar actors (Boschma and Frenken, 2010). This attractiveness is of a non-dyadic nature and refers to relational states of actor proximity or similarity.

The theoretical concept behind this reasoning is homophily, stating that connections are established based on the similarity of the actors involved (Rivera et al., 2010; McPherson et al., 2001; Freeman, 1996). Tie formation based on similarities within the network can be based on restricted opportunities to connect to dissimilar actors induced by the group to which an actor belongs, and by preferences to connect to similar partners (McPherson and Smith-Lovin, 1987). A high level of similarity among the actors of a (sub)network promotes mutual understanding and thus, influences the frequency and intensity of communication and interaction as well as the joint use of knowledge and other resources. Hence, interaction within homogeneous (sub)networks is subject to a self-reinforcing process generated within the network (Rogers, 1995). In order to benefit from the frequent interaction in homogeneous (sub)networks, the actors build up new ties to actors showing similar characteristics.

The relevance of similarity for network formation has found empirical support on various levels such as on the individual level in partnering choices (e.g. Bozeman and Corley, 2004; Ruef et al.,

2003; McPherson et al., 2001; McPherson and Smith-Lovin, 1987); for inter-organizational alliances in German stock photography organizational similarity in terms of firm characteristics appear as a weak explanation (Glückler, 2010) whereas in the biotechnology industry alliance formation is related to similarities in the social status of the management and the role firms play in the industry (Kim and Higgins, 2007).

Related to our country level analysis it has been shown that similarity with respect to the level of economic development, culture, and the size and specialization of the national scientific infrastructure influence the choice of collaboration partners (Zitt et al., 2000; Luukkonen et al., 1992). Since mutual understanding based on economic, cultural and scientific similarities may drive the dynamics of cross-country collaboration we formulate Hypothesis 2 as follows:

Hypothesis 2. The degrees of economic, cultural, and scientific similarities between two countries are positively related to changes in the number of cross-country co-author collaborations between them.

2.2.3. Multi-connectivity

With the concept of connectedness we address a quite narrow dimension of attractiveness based on dyadic relationships and with similarity attractiveness is seen in a non-dyadic but rather broad sense. With our third concept, multi-connectivity, we address the indirect relations between actors that potentially will cooperate.

Coleman (1990, 1988) has argued that closely connected networks increase cooperation and trust among the actors and thus knowledge exchange. Hence, the actors in the network have also incentives to build up linkages to those actors they are already indirectly connected with (Cowan and Jonard, 2007). Building upon these arguments, Powell et al. (2005) suggest that multi-connectivity can be the underlying mechanism that leads to the emergence of cohesive network structures. Accordingly, actors who are more diversely linked with each other in $t-1$ are more likely to form a new tie in t than pairs of actors with less diverse indirect connections. In this sense, it is multi-connectivity (as measured by the number of indirect ties between two potential partners) that drives tie formation – a dynamics of closure and then of clustering Rivera et al. (2010).

The impact of multi-connectivity on tie formation finds empirical evidence in the case of strategic alliances in the German stock photography market (Glückler, 2010) and in life sciences between different types of organizations (Powell et al., 2005). In the former study two actors are more likely to engage in a partnership if they are connected via third parties and in the latter study this likelihood is higher for actors that are more diversely connected to each other in the previous period.

Against this background, we analyze whether multi-connectivity is related to the co-authorship dynamics on the country level. In line with the arguments outlined above we suggest a relationship between multi-connectivity and cross-country collaboration dynamics as follows:

Hypothesis 3. The degree of multi-connectivity between two countries is positively related to changes in the number of cross-country co-author collaborations between them.

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, the development of technologies (Mina et al., 2007; Verspagen, 2007), and the importance of the structure of collaborative relations for innovative activities (Ahuja, 2000; Baum et al., 2000; Burt, 1992). 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 one another. 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 collaboration in the pharmaceutical industry, we refer to countries as 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 collaboration networks in different therapeutic areas, we compute several descriptive statistics. The number of countries describes the count 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 one another. 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 (edges) of a graph divided by the maximal possible number of edges:

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

where g is the group size, i.e. the number of nodes in the graph, and $d(n_i)$ 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 what 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 have used the degree of an node as a 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 completely isolated.

Following 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 observed variation in the actor's degree indices (numerator) divided by the maximum possible variation (denominator). The denominator can be expressed directly by $(g-1)(g-2)$ (cf. Freeman, 1979), leading to:

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

Eq. (6) gives an index of the degree of centralization of the network's set of actors. 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 a 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 that 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 extent 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 edges 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 total number of shortest paths linking actors j and k containing actor i. The probability that two actors, j and k, are linked by an distinct actor i is given by $g_{jk}(n_i)/g_{jk}$. The index $C_B(n_i)$, which accounts for i's betweenness with respect to all actors j and k, can be standardized so that it takes values between 0 and 1 and can be compared between among different actors and networks:

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

The application of group betweenness centralization measures allows us to compare different networks with respect to the variation of the actors' betweenness. According to Freeman (1977, 1979), we can express the group betweenness centralization index as:

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

In the numerator, $C_B(n_i)$ represents the actor betweenness index and $C_B(n^*)$ its 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 edges and all nodes are distinct. The length of a path is its number of edges. The average path length is defined as the average number of edges along the shortest paths between all nodes of the network:

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

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 as to whether two actors connected to a third one interact among one another, 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 triads N_Δ and the number of connected triples N_3 in the graph.

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

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

The point connectivity and the geodesic count can be used to assess the connectivity of nodes within a network. More precisely, the point connectivity of each node in the network is defined as the minimum number of other nodes that have to be removed from the network in order to disconnect a pair of nodes. The geodesic count refers to the number of shortest paths connecting two nodes in the network.

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:

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

The elements $y_{i,j}$ of the matrix Y equal zero if there is no relation between actor i and actor j and are equal to any other value otherwise. Thus, the values of $y_{i,j}$ indicate the strength of the relation between both actors. 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):

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

Here, 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 amount of collaboration. 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 require different techniques to examine the coefficients and particularly their the significance, since the assumptions of the standard OLS model are usually violated, e.g. by structural autocorrelation, which frequently appears either in rows or columns of the network matrix (Krackhardt, 1987). Thus, conventional test statistics may provide misleading standard errors and significance levels. The multiple regression quadratic assignment procedure (MRQAP) has been found to be an appropriate method to derive more correct inferences concerning the significance of the model's coefficients (Hubert, 1987). This procedure provides a general, permutation-based, non-parametric test of the significant relation of two structures (see among others Hubert and Schultz, 1976; Mantel, 1967). The general idea of MRQAP is to generate the reference distribution by random permutation of original data matrix' rows and columns against which the coefficients are compared. 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 MRQAP procedure has been found to be quite robust against autocorrelation encountered in network data. We use the double semi-partialing method (DSP) proposed by Dekker et al. (2007, 2003), since it provides a version of the MRQAP procedure that is robust against multicollinearity and other conditions such as skewness of the data. MRQAP models require a relatively large number of random permutations. In our study, we use 10,000 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 a 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.[2] 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 that correspond to a system of an organism or a general disease group.[3] 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 Science databases (WoS). The WoS consist of seven databases containing information gathered from an extensive number of journals, books, book series, reports and conferences. Among these databases, the most important is the Science Citation Index Expanded (SCI), a multidisciplinary index of more than 6500 scientific journals, covering 150 scientific disciplines. The SCI covers, among others, the scientific fields of biochemistry, medicine and pharmacology which are of particular interest for our study. The WoS includes information concerning the scientific publications themselves, such as the title, the year of publication, the journal, cited references, a categorization of the research fields, to which a publication can be assigned, and further bibliographic information. In addition to this information, the WoS reports for most articles the authors' affiliations and their addresses 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 our keyword list in their title have been included in our dataset. In order to refine the results, we only take into account publications included in categories related to pharmaceutical research. 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.[4] 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 113,057 articles. We further restrict our sample to all articles that contain information concerning the authors' affiliations. In total, our sample consists of the 111,096 journal articles. In order to analyze the development of cross-country scientific collaboration over time, we distinguish three sub periods, 1998–2000, 2002–2004, and 2006–2008. 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.

We use the reported author affiliations to extract information concerning the countries of origin of the scientific articles in our dataset. Consequently, our networks encompass all countries that have been assigned to at least one publication in the respective therapeutic area and period. Country level publication data is matched with World Bank income groups in order to have

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

[3] Table 6 (see Appendix A.1.) provides an overview of the therapeutic areas included in the dataset.

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

some information concerning the wealth level of the countries in our sample. Articles in the categories "Biochemistry & Molecular Biology" and "Biotechnology and Applied Microbiology" are regarded as biotechnology publications. The CHI classification of journals (Hamilton, 2003) gives us the opportunity to classify each article according to the type of research prevalent in the journal, in which it is published. The application of this classification scheme enables us to distinguish "clinical observation", "clinical mix", "clinical investigation", and "basic biomedical research" publications. We employ the CEPII (Centre d'Études Prospectives et d'Information Internationales) database on distance measures in order to get information concerning language similarities among countries (Mayer and Zignago, 2006). We use the World Bank Science & Technology database in order to get information concerning the overall scientific output of country in terms of journal articles in scientific and technical journals.

Publication data provide 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 when analyzing co-publication data. The most important are that research does not necessarily lead to publication, co-authorship may only partly capture scientific collaboration, the impact of publications differs considerably and publication habits differ among scientific disciplines. Publication databases may be biased towards English language publications and journals published in industrialized countries. 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; Hoekman et al., 2009 for a discussion).

4. International research networks

4.1. Network descriptives

In this section, we employ social network analysis to analyze differences in the cross-country collaboration patterns in pharmaceutical research in various therapeutic areas. We use the igraph package by Garbor Csardi and netmodels package by Domingo Vargas for R statistical software to calculate descriptive network statistics.

We start our analysis taking into account all journal publications in the respective therapeutic areas and periods. 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 cross-country research community and the relative size of the largest component, i.e. the largest group of connected countries, increase in almost all therapeutic areas, from the first to the third period. This corresponds to a decrease in the share of isolated countries, which do not collaborate with other countries. However, their absolute number increases in eight 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 countries to which a country is connected 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.

The degree centralization measure equals values above 0.4 in most networks over all three periods, indicating that the number of linkages is quite dispersed among countries in the majority of the analyzed networks. This finding indicates that some countries collaborate more than others. All betweenness centralization measures are below 0.42, which indicates some dispersion of these measures among the countries 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 cross-country collaboration.

We analyze which countries are located at central network positions in Table 7 (see Appendix A.2.). In all therapeutic areas high income OECD countries account for a substantial but declining share of the countries in the networks. This development can be explained by the entry of non-OECD countries. The connectivity of OECD and non-OECD countries in terms of the mean degree of countries increases over time. However, the mean degree of OECD countries is considerably larger compared to the mean degree of non-OECD countries in most therapeutic areas. We compute the average betweenness centrality scores for OECD and non-OECD countries to obtain some further information concerning the position of these country groups in the network. The results reveal that the average betweenness centrality scores increase over time for both groups of countries while the scores of OECD countries remain considerably larger than those of non-OECD countries. Hence, we find that particularly OECD countries are located at central positions within the network of cross-country scientific collaborations. The results stay qualitatively similar if we drop the United States as the country with the most publications and collaborations from the analysis.

In further steps, we restricted our analysis to basic research and biotechnology publications in order to examine whether the trend towards increasing collaboration described earlier can be found in this subfield as well. The number of countries involved in these types of research is, in general, somewhat lower compared to the complete networks. Nevertheless, the number of involved countries increases over time in most networks and the countries increase collaboration among one another. As in the case of the complete networks, high income OECD countries can be found in the center whereas developing and newly industrializing countries can be found in peripheral positions of the network. Consequently, the cross-country research network in the fields of basic and biotechnology research show similar patterns as the networks, including all journal articles in the respective therapeutic areas. In order to ensure that the increasing cross-country collaboration is not driven by an expanding number of journals, we restrict our sample to those journals that have been included in the WoS prior to 1998 according to the CHI classification. The results for this subsample are in line with the original analysis. The analysis of weighted instead of binary networks reveals a trend towards increasing collaboration and cohesion. The mean degree, the average collaboration intensity, and the clustering coefficient are increasing over time in all therapeutic areas. Again, high income OECD countries can be found in central positions within the networks.

4.2. Entry and exit

In the previous section, network statistics seem to indicate intensified collaboration across countries in almost all therapeutic areas. We find that an increasing number of countries are engaged in collaborative pharmaceutical research across borders. In this section, we analyze the number of entries, exits and persistently

Table 1
Network descriptive statistics.

Therapeutic area ID	Period	Number of countries	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
All	1	136	7	130	0.956	6	0.044	0.107	14.397	0.576	0.221	2.070	0.427
All	2	141	9	133	0.943	8	0.057	0.119	16.723	0.582	0.195	2.045	0.487
All	3	154	1	154	1.000	0	0.000	0.136	20.779	0.597	0.167	2.068	0.499
1	1	73	11	63	0.863	10	0.137	0.109	7.863	0.530	0.244	2.091	0.449
1	2	84	9	76	0.905	8	0.095	0.120	9.929	0.556	0.221	2.098	0.483
1	3	101	7	95	0.941	6	0.059	0.127	12.673	0.554	0.212	2.092	0.492
2	1	73	15	58	0.795	13	0.178	0.091	6.548	0.449	0.166	2.184	0.443
2	2	84	15	70	0.833	14	0.167	0.082	6.786	0.422	0.157	2.309	0.443
2	3	89	15	75	0.843	14	0.157	0.118	10.382	0.402	0.122	2.226	0.535
3	1	56	7	50	0.893	6	0.107	0.141	7.750	0.495	0.259	2.024	0.512
3	2	68	9	60	0.882	8	0.118	0.123	8.235	0.596	0.286	2.023	0.453
3	3	79	10	70	0.886	9	0.114	0.127	9.899	0.527	0.207	2.082	0.500
4	1	31	20	5	0.161	16	0.516	0.024	0.710	0.117	0.013	1.565	0.000
4	2	32	17	16	0.500	16	0.500	0.054	1.688	0.183	0.113	2.358	0.510
4	3	35	18	18	0.514	17	0.486	0.049	1.657	0.260	0.153	2.418	0.351
6	1	48	16	33	0.688	15	0.313	0.058	2.708	0.406	0.242	2.388	0.297
6	2	54	12	43	0.796	11	0.204	0.084	4.444	0.481	0.307	2.174	0.353
6	3	69	16	53	0.768	14	0.203	0.067	4.580	0.491	0.254	2.146	0.337
7	1	67	16	51	0.761	14	0.209	0.071	4.687	0.364	0.200	2.460	0.451
7	2	68	14	55	0.809	13	0.191	0.083	5.559	0.499	0.242	2.221	0.432
7	3	77	14	64	0.831	13	0.169	0.096	7.325	0.482	0.175	2.185	0.430
8	1	42	14	29	0.690	13	0.310	0.057	2.333	0.401	0.326	2.495	0.282
8	2	44	14	30	0.682	12	0.273	0.056	2.409	0.429	0.319	2.326	0.274
8	3	55	12	44	0.800	11	0.200	0.071	3.855	0.464	0.337	2.314	0.347
9	1	59	14	44	0.746	12	0.203	0.061	3.525	0.276	0.190	2.526	0.383
9	2	55	14	41	0.745	12	0.218	0.065	3.491	0.433	0.265	2.352	0.305
9	3	63	14	50	0.794	13	0.206	0.084	5.206	0.513	0.345	2.287	0.528
10	1	24	11	14	0.583	10	0.417	0.098	2.250	0.415	0.168	1.824	0.425
10	2	28	11	18	0.643	10	0.357	0.074	2.000	0.439	0.271	2.078	0.250
10	3	38	14	25	0.658	13	0.342	0.077	2.842	0.318	0.165	2.307	0.414
11	1	59	15	44	0.746	13	0.220	0.063	3.627	0.399	0.261	2.317	0.314
11	2	64	12	53	0.828	11	0.172	0.082	5.156	0.473	0.315	2.294	0.384
11	3	72	12	61	0.847	11	0.153	0.129	9.194	0.446	0.154	2.086	0.515
12	1	58	8	50	0.862	6	0.103	0.084	4.793	0.603	0.419	2.151	0.282
12	2	56	11	45	0.804	9	0.161	0.110	6.071	0.489	0.262	2.053	0.475
12	3	72	12	61	0.847	11	0.153	0.103	7.306	0.474	0.197	2.168	0.484
13	1	116	13	104	0.897	12	0.103	0.080	9.224	0.458	0.186	2.189	0.359
13	2	121	7	115	0.950	6	0.050	0.109	13.091	0.508	0.206	2.154	0.467
13	3	132	4	129	0.977	3	0.023	0.111	14.576	0.585	0.188	2.104	0.399
15	1	50	15	34	0.680	12	0.240	0.080	3.920	0.384	0.152	2.062	0.352
15	2	52	11	42	0.808	10	0.192	0.102	5.192	0.465	0.194	2.156	0.399
15	3	65	16	50	0.769	15	0.231	0.072	4.585	0.474	0.253	2.274	0.385
16	1	45	12	27	0.600	8	0.178	0.084	3.689	0.293	0.109	2.000	0.433
16	2	44	13	31	0.705	11	0.250	0.122	5.227	0.360	0.159	2.026	0.505
16	3	54	15	39	0.722	13	0.241	0.091	4.815	0.376	0.181	2.082	0.423
17	1	67	7	61	0.910	6	0.090	0.112	7.373	0.447	0.308	2.268	0.465
17	2	62	11	52	0.839	10	0.161	0.106	6.484	0.483	0.203	2.127	0.436
17	3	77	10	68	0.883	9	0.117	0.095	7.195	0.484	0.211	2.277	0.403

contributing countries. In doing so, we calculate the mean degree, i.e. the average number of connections a country has, for the three subgroups mentioned. The connectivity of countries within the network may be associated with their research performance and their decision to leave the network. Based on evidence on the 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 countries (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 13 out of 15 therapeutic areas, at least ten countries enter, and in six therapeutic areas, the number of exits is at least ten in the period 2002 to 2004. The number of entering countries exceeds the number of exits in eleven therapeutic areas. In the third period, we find positive 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. The positive net entry in most therapeutic areas, particularly in the third period, suggests, again, that scientific collaboration in pharmaceuticals has become more international. Moreover, entries and exits give us some hint that there is some dynamic in the formation and break-up of ties within the networks.

With respect to the mean degree of each subgroup, entering, exiting and permanent countries (incumbents), we find considerable differences in all therapeutic areas among these groups. Incumbents are connected to a by far higher number of other countries than entering and exiting countries.[5] This finding is prevalent for entries and exits from the first to the second and from the second to the third period. With respect to the exiting countries, we interpret this as a hint that these countries left the cross-country research network because of a relatively weak position in the respective field in terms of international contacts. For countries entering in the third period, we find, on average, a higher number of connections than for exiting countries. Nevertheless, entering countries are far less connected than incumbents. The latter increase their average number of collaborative ties in 13 out of 15 therapeutic areas. This finding indicates that these countries increasingly engage in cross-country research collaboration.

The networks taking into account only basic research, biotechnology and articles published in journals included in the WoS prior to 1998 show very similar patterns of entry and exit. Again, the number of entries and exits is considerable and exiting countries are far less connected than incumbents.

5. Empirical results network regressions

5.1. Variables

We present here an overview of the variables and controls used in our network regression models in Table 3. Descriptive statistics are presented in Table 4. The dependent variable is the change in the number of total collaboration between two countries between period $t-1$ and period t. More precisely, we calculate the amount of collaboration for each pair of countries in period t and subtract the amount of collaboration in period $t-1$. This variable captures changes in the intensity of collaboration activities, the formation of collaboration ties between previously unconnected countries as well as the break-up of any previously existing ties. Collaboration activities are measured via co-publication activities. The number of co-publications between each pair of countries is calculated based on author affiliations. We use full counting, which leads to a

[5] The mean degree for exiting countries refers to the previous period in which they were part of the network.

Table 2
Entries, exits and incumbent countries.

	1998–2000	2002–2004						2006–2008					
Therapeutic area ID (incumbents in 2002–2004)	Mean degree of incumbents	Entries	Mean degree entries	Exits	Mean degree exits	Incumbents	Mean degree incumbents	Entries	Mean degree entries	Exits	Mean degree exits	Incumbents	Mean degree incumbents
All	15.983	20	2.400	15	1.600	121	19.091	22	2.273	9	1.444	132	23.864
1	8.859	20	1.100	9	0.778	64	12.688	26	3.000	9	1.222	75	16.027
2	6.803	18	1.056	7	4.143	66	8.348	16	1.375	11	0.909	73	12.356
3	8.646	20	2.200	8	2.375	48	10.750	22	3.455	11	1.727	57	12.386
4	0.846	7	0.143	6	0.167	25	2.120	10	0.300	7	0.286	25	2.200
6	3.125	14	1.357	8	0.625	40	5.525	24	0.958	9	2.222	45	6.511
7	5.453	15	1.267	14	1.786	53	6.774	19	2.053	10	1.600	58	9.052
8	2.667	11	0.273	9	1.111	33	3.121	16	1.063	5	0.200	39	5.000
9	4.488	14	0.714	18	1.333	41	4.439	19	0.947	11	1.091	44	7.045
10	3.000	10	1.200	6	0.000	18	2.444	18	1.111	8	0.875	20	4.400
11	4.098	13	1.692	8	0.625	51	6.039	17	2.765	9	1.444	55	11.182
12	5.435	10	1.000	12	2.333	46	7.174	23	2.348	7	0.429	49	9.633
13	9.943	15	4.800	10	1.600	106	14.264	20	1.600	9	3.778	112	16.893
15	4.700	12	1.583	10	0.800	40	6.275	19	1.158	6	2.667	46	6.000
16	4.111	8	0.500	9	2.000	36	6.278	13	1.000	3	0.333	41	6.024
17	8.875	14	0.857	19	3.579	48	8.125	24	2.708	9	1.222	53	9.226

Table 3
Overview of variables.

Dependent variable		
ΔCollaboration		Change in the number of collaborations between two countries from period $t-1$ to t
Independent variables		
Diff_DegreeCentrality	Connectedness	Difference in the degree centrality between two countries lagged by one period
Collaboration	Connectedness	Amount of collaboration between two countries lagged by one period
IncomeSimilarity	Similarity	Dummy indicating whether two countries belong to the same income group
LanguageSimilarity	Similarity	Dummy indicating if at least 9% of the population in both countries speak the same language
Diff_ResearchStrength	Similarity	Difference in the total number of science and technology articles per one million inhabitants between two countries lagged by one period
PointConnectivity	Multi-connectivity	Number of other countries that have to be removed in order to disconnect two countries lagged by one period
GeodesicCount	Multi-connectivity	Number of shortest paths between two countries lagged by one period

Table 4
Descriptive statistics explanatory variables MRQAP regression.

Variable	Period	Observations	Mean	Std. Dev.	Min	Max
ΔCollaboration	2	13110	0.688	11.788	−217.000	721.000
	3	15500	1.427	22.573	−141.000	1347.000
Diff_DegreeCentrality	2	13110	0.065	0.114	0.000	0.674
	3	15500	0.070	0.116	0.000	0.693
Collaboration	2	13110	1.430	21.751	0.000	1368.000
	3	15500	1.798	26.759	0.000	1371.000
IncomeSimilarity	2	13110	0.097	0.296	0.000	1.000
	3	15500	0.093	0.291	0.000	1.000
LanguageSimilarity	2	13110	0.061	0.240	0.000	1.000
	3	15500	0.058	0.234	0.000	1.000
Diff_ResearchStrength	2	13110	115.374	241.692	0.000	1151.029
	3	15500	113.039	238.097	0.000	1121.178
PointConnectivity	2	13110	3.601	6.678	0.000	67.000
	3	15500	3.984	7.878	0.000	75.000
GeodesicCount	2	13110	1.883	4.141	0.000	94.000
	3	15500	2.009	4.726	0.000	90.000

co-publication count of one for each pair of countries involved in a publication. Since co-publications represent undirected links, each pair of countries is included only once in a specific period and therapeutic area.

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. With respect to connectedness we distinguish differences in the degree of connectedness in the network between two countries and the degree by which two countries are connected. For the former which draws on the network embeddedness of a country we follow Glückler (2010) and use absolute differences in countries' degree centrality scores lagged by one period as a proxy for connectedness (*Diff_DegreeCentrality*). This measure refers to differences in the visibility of countries in the research network. For the latter which addresses the direct connectedness of two countries, the number of prior ties has been used as an indicator for an accumulative advantage based on previous connectedness (cf. Powell et al., 2005). Therefore, we include in our analysis the number of previous collaboration among two countries lagged by one period (*Collaboration*).

With respect to similarities between countries we distinguish three dimensions. A first one is reflected by the variable *IncomeSimilarity* indicating whether two collaborating countries belong to the same World Bank income group, i.e. they have comparable wealth levels. We use language similarities among countries as a second measure for similarity. More precisely, *LanguageSimilarity* equals 1 if at least 9% of the population speaks the same language. Third, *Diff_ResearchStrength* accounts for similarities in the research strength of two countries. To obtain this measure, we calculate the difference in the total number of articles in science and technology journals per one million inhabitants lagged by one period.[6] Multi-connectivity 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 (prospectively) collaborating countries. Moreover, we use the number of shortest paths between two countries in the network with a lag of one period (*GeodesicCount*) as a further proxy for multi-connectivity (cf. Glückler, 2010). These two measures, although addressing the same principle dimension, focus on different aspects of multi-connectivity; *PointConnectivity* looks at the number of different in-between countries whereas *GeodesicCount* counts the number of shortest paths. For the latter, it has to be recalled that values greater 1 imply no direct connections between two countries.

5.2. Regression results

In Table 5, we present the results of our regression analysis on the formation and break-up of network ties as measured

[6] According to the World Bank Science & Technology database this measure includes articles from a broad variety of fields including physics, biology, chemistry, mathematics, clinical medicine, biomedical research, engineering and technology, and earth and space sciences.

Table 5
Network regression.

	Period 2				Period 3			
	(1)		(2)		(3)		(4)	
	Estimate	Pr(≥|b|)	Estimate	Pr(≥|b|)	Estimate	Pr(≥|b|)	Estimate	Pr(≥|b|)
Dependent variable: ΔCollaboration								
Diff_DegreeCentrality	5.3585	0.0127	9.9599	0.0003	2.3331	0.1273	5.6178	0.0242
Collaboration	0.2066	0.0000	0.1986	0.0000	0.6050	0.0000	0.6017	0.0000
IncomeSimilarity	0.9575	0.1141	0.4890	0.3677	−0.6612	0.2279	−1.1592	0.0907
LanguageSimilarity	−1.2698	0.0955	−1.4303	0.0772	1.1834	0.0923	1.2593	0.0886
Diff_ResearchStrength			−0.0045	0.0010			−0.0026	0.0425
PointConnectivity	0.3179	0.0000	0.3741	0.0000	0.1841	0.0000	0.2277	0.0001
GeodesicCount	−0.2228	0.0032	−0.2035	0.0049	−0.0665	0.0678	−0.0710	0.0698
Intercept	−0.5978	0.1333	−0.5206	0.1656	−0.3934	0.2237	−0.3755	0.2413
Residual standard error	13.17		13.04		15.37		15.69	
F-statistic (*p*-value)	323.5	0.0000	274.7	0.0000	1672	0.0000	1377	0.0000
Adjusted *R*-squared	0.2279		0.2345		0.5641		0.5657	

Nullhypothesis: MRQAP with DSP and 10,000 permutations.

by the change in the number of total collaboration between two countries. We perform our analysis on the aggregated or pooled level, i.e. we do not distinguish among the different therapeutic areas and we concentrate on countries that are members of the network in periods $t-1$ and t, the incumbent countries. For that sample we can use most of our independent variables as lagged by one period allowing us an analysis that goes beyond pure correlation and delivers some insight into the factors and mechanisms driving the formation and the break-up of ties within a network.[7] Network correlations of the independent variables can be found in Tables 8 and 9 (see Appendix A.3.). Since the correlation between *Diff_ResearchStrength* and *Diff_DegreeCentrality* is quite high, we present separate models that include and do not include *Diff_ResearchStrength*. Nevertheless, the MRQAP procedure with double semi-partialing permutation is supposed to be quite robust against multi-collinearity (Dekker et al., 2007, 2003). Results not presented in the paper are available upon request.

With respect to Hypothesis 1a and relative connectedness as a driver of tie formation, we find a positive and significant coefficient for *Diff_DegreeCentrality* in period 2 and when we introduce the variable together with *Diff_ResearchStrength* in period 3. For the second period, this indicates a positive relation between differences in the degree centrality of countries lagged by one period and changes in the intensity of collaboration. Hence, a larger difference of countries in terms of their relative connectedness promotes further collaboration ties among them implying that new or intensified collaboration will take place between unequally embedded countries whereas the break-up of ties will be observed more likely between more equally embedded countries. Since the respective coefficient is not significant in model (3) for the third period, results are not robust with respect to time. The indicator for bilateral connectedness proxied by *Collaboration*, shows a positive and significant association towards changes in the amount of collaboration between two countries in subsequent periods.[8] This result can be interpreted as a hint that countries' bilateral connectedness measured by previous collaboration can lead to an accumulative advantage based on connectedness which induces further collaboration activities as suggested by Hypothesis 1b. Put differently, a joint collaboration experience may lead to a self-reinforcing process of intensified collaboration in which countries that have been well connected in previous periods form more intensely new ties among each other. Taken our two results for connectedness together, we find evidence for the existence of a "rich-get-richer" phenomenon in which the well-connected countries can further increase their connectivity. These findings are to a wide extent in line with the theoretical predictions underlying Hypotheses 1a and 1b.

Similarity dimensions are at the core of Hypothesis 2. Similarity in terms of countries being in the same income group (*IncomeSimilarity*) is, with an exception in model (4), not significantly related to the formation and break-up of research collaboration. Hence, our results do not suggest that either similarities or differences in terms of income groups are robustly associated with changes in the amount of collaboration. With respect to language similarities (*LanguageSimilarity*), we find a weakly significant negative relationship of the same language spoken in two countries and tie formation and break-up in period 2. However, in period 3, we find a weakly significant positive association. Consequently, our results do not suggest that similarity in terms of language among countries has a robust, clear-cut relationship to changes in the amount of collaboration at the country-level. Differences in the total number of science and technology journal articles per one million inhabitants as an indicator for differences in the strength of countries' science systems (*Diff_ResearchStrength*) are negatively and significantly related to the formation and break-up of research collaboration. This finding suggests that a higher similarity of countries in terms of their population adjusted scientific output leads to increasing interactions among these countries. Taken these three results together, we find that some particular forms of similarity among countries affect the dynamics of international collaboration activities in pharmaceutical research while others do not. Neither countries' economic nor language similarities (taking also into account that the *lingua franca* in science is English) are linked to the dynamics of international collaboration activities in pharmaceutical research. In contrast to these findings, the general research performance of the respective national research and innovation systems proxied by the total number of science and technology articles is positively linked to changes in the intensity of scientific collaborations across countries as suggested by theoretical concept of similarity based network dynamics. The latter finding supports Hypothesis 2 and indicates that a similar level of research intensity promotes mutual

[7] Ideally, tie formation of entering countries would give some insights concerning the mechanisms driving the dynamics of the network. However, the problem with this approach is that lagged variables for entrants are not available, which makes it hard to identify the mechanisms at work with more sophisticated methods. Therefore, we concentrate on tie formation and break up among incumbents for which lagged variables are available.

[8] *Collaboration* is the main source of differences in the adjusted *R*-squared, since it contributes much less to this measure in period 2 compared to period 3. We account for the possibility that the *Collaboration* variable dominates our results. More precisely, we run additional regressions without the *Collaboration* variable using the number of collaborations among countries that did not collaborate in the previous period as dependent variable. The results are qualitatively similar to those obtained in the original analysis.

understanding and provide the basis for cross-country research collaboration.

Via Hypothesis 3 we analyze whether multi-connectivity, accounting for countries' bilateral network attractiveness, is suitable to explain changes in the amount of research collaboration on the country level and find a positive and significant coefficient for *PointConnectivity*. This finding suggests that changes in the intensity of collaboration are positively related to the number of countries that indirectly connect two other countries. Put differently, the intensity of collaboration may change due to knowledge flows the partners receive through other collaboration. The coefficient for *GeodesicCount*, i.e. the number of shortest paths, has a significantly negative sign in both periods. The sign of the coefficient is rather intuitive, since a high number of shortest paths indicates that there has been no direct interaction among two countries, which is also shown by the slightly negative correlation between *Collaboration* and *GeodesicCount* in Tables 8 and 9 (see Appendix A.3.). Hence, our results suggest that multiple shortest paths as a proxy for multi-connectivity are negatively associated with tie formation and break up in both periods. Taken these two results together we find good support for Hypothesis 3. As the theoretical concept of multi-connectivity would predict, there appears to be a positive relation between indirect ties of two countries and changes in their cross-country research collaborations which is only reduced when the two countries have been not or only weakly connected directly.

Our results for all three hypothesis stay qualitatively similar if we restrict our sample to collaboration in the fields of basic and biotechnology research, as well as to those journals included in the WoS prior to 1998. Moreover, we obtain similar results when we additionally control for isolated countries that did not collaborate in the previous period.

6. Conclusion

Literature suggests that knowledge production and scientific research are increasingly 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 pharmaceutical research collaboration networks at the country level in different therapeutic areas. Our empirical analysis is based on a unique dataset of journal publications related to pharmaceutical research. By means of social network analysis, we find that the cross-country research networks expand over time in almost all therapeutic areas. More specifically, the number of countries involved and their connectivity increases in most therapeutic areas. High income OECD countries are located in the core of the cross-country research networks. This pattern remains rather stable over time.

In order to assess which mechanisms, namely connectedness, similarity, or multi-connectivity, drive the endogenous network dynamics, we employ multiple regression analysis for dyadic data. More precisely, we use the MRQAP procedure with double semi-partialing permutation. Our regression results reveal a positive association between the connectedness of two countries as measured by the amount of previous research collaboration between two countries and the change in their amount of collaboration, indicating an accumulative advantage based on the connectedness of countries. Differences in countries' degree centrality as proxy for connectedness in terms of network embeddedness show no robust significant relation to changes in the collaboration intensity. Our results do not allow for a clear-cut conclusion whether similarities in terms of per capita income and language are driving in the change in cross-country collaboration. Differences in the strength of countries' research systems are, however, negatively related to changes in the collaboration intensity. Multi-connectivity in terms of different countries connecting two other countries is positively related, whereas the number of shortest paths shows a negative association with changes in the amount of collaboration.

Our empirical results are in accordance with literature suggesting the growing amount of collaborative work on the national and international level (e.g. Mattsson et al., 2008; Adams et al., 2005). Our measures of the network structures and the number of entries and exits reveal that the cross-country networks are changing over time. As to the driving mechanisms behind tie formation and break-up we refer to the literature addressing a couple of different sources. There has been empirical evidence for connectedness (e.g. Orsenigo et al., 1998), similarity (e.g. Glückler, 2010), and multi-connectivity (e.g. Powell et al., 2005) being the mechanism of tie formation in different real world networks. Our regression results indicate that these different mechanisms influence the formation and break-up of ties between "collaborating countries". Hence, the internationalization of research collaboration is influenced by the relative and bilateral connectedness of countries indicating the presence of a "rich-get-richer" phenomenon in which, as in the Barabási and Albert (1999) model, already well connected countries increase their connectivity even further. Additionally, we find that similarities in terms of countries' research strength supports the theoretical argument and the empirical evidence of connections being preferably established between similar actors (McPherson et al., 2001). These mechanisms may lead to an increasing connectivity within the network that implies the potential for knowledge transfer. In order to secure this potential for (joint) knowledge generation and recombination, countries engage in the formation of multiple paths through to their collaboration partners as suggested by the multi-connectivity approach (Powell et al., 2005).

Our results concerning the structure and the dynamics of the international research networks in pharmaceuticals may have important policy implications. Countries that are not part of the center of the network may have difficulties to get access to the sources of knowledge required for successful generation and usage of pharmaceutical innovations. Therefore, policy makers may support the access to the relevant knowledge sources by supporting international collaboration between institutions from countries located in the center of the pharmaceutical research networks and countries located in the periphery. Countries located in the networks' periphery have typically few connections and are not linked through multiple other countries to their collaboration partners. Hence, the dynamics of the networks may weaken their position even further by generating more new collaborations between countries that belong to the core of the networks. Active policy support is to be seen as an "entry ticket" into international collaboration, a "ticket" that may help to overcome the liabilities of unconnectedness of countries that are not part of the core of the network. When this first step is done, further advances and steps of embedding into international collaboration networks may follow rather automatically. Then, the experience gathered by past collaborations and by building multiple connections to other collaborating countries will be an important driver of further integration into the cross-country networks. Policy makers in countries that are not part of the core of the network may enforce such kind of development by supporting and establishing direct linkages to countries in the core or indirectly by investing more into their national research and innovation system.

Since our investigation is restricted to pharmaceuticals, future research may focus on the development of cross-country research collaboration in different industries. The pharmaceutical industry may provide an exceptional case due its pronounced scientific foundation and the importance of (international) collaboration networks. Hence, country level factors may have different effects for the structure and the dynamics of international research networks in other fields.

Acknowledgements

We thank the German Research Foundation (DFG) for financial support. We are grateful to Abiodun Egbetokun, the participants of the 4th ZEW Conference on Economics of Innovation and Patenting & Finale Conference COST Network STRIKE 2011, the participants of the 5th Jena Summer Academy on "Innovation and Uncertainty", and two anonymous referees for useful comments, expressed interest and concerns. The usual caveats apply.

Appendix A.

A.1. List of therapeutic areas

A.2. Network position of OECD countries

A.3. Network correlations

Table 6
List of therapeutic areas.

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 7
Network descriptive statistics OECD countries.

Therapeutic area ID	Period	Share OECD countries in network	Mean degree OECD countries	Mean degree non-OECD countries	Mean betweenness centrality OECD countries	Mean betweenness centrality non-OECD countries
All	1	0.199	26.111	17.679	48.259	21.945
All	2	0.191	30.889	19.079	57.630	19.553
All	3	0.175	38.519	22.024	63.704	32.323
1	1	0.370	13.852	8.457	17.963	6.326
1	2	0.321	15.148	10.474	24.889	7.842
1	3	0.267	16.370	12.757	40.704	14.581
2	1	0.342	12.720	8.083	13.760	3.521
2	2	0.298	14.480	8.017	23.680	4.610
2	3	0.281	16.320	12.328	30.840	11.875
3	1	0.446	10.360	10.581	16.080	2.903
3	2	0.397	12.704	10.561	13.519	4.854
3	3	0.342	13.222	13.038	19.444	7.346
4	1	0.645	2.000	1.182	1.350	1.000
4	2	0.688	3.409	1.100	2.364	1.000
4	3	0.571	3.800	1.467	3.750	1.000
6	1	0.479	5.217	2.160	5.217	1.000
6	2	0.463	9.120	5.862	11.240	1.690
6	3	0.348	9.083	5.978	10.625	2.000
7	1	0.388	9.115	5.683	9.154	2.756
7	2	0.368	9.720	6.419	11.520	1.860
7	3	0.325	13.080	10.654	20.560	4.404
8	1	0.476	4.750	2.182	4.300	1.227
8	2	0.477	4.571	1.913	3.857	1.435
8	3	0.400	8.182	4.000	8.227	2.242
9	1	0.407	6.333	4.543	6.833	2.171
9	2	0.418	7.000	4.281	8.043	2.156
9	3	0.413	8.769	4.405	10.115	3.000
10	1	0.625	3.933	2.222	2.733	1.222
10	2	0.464	3.462	2.333	2.462	1.133
10	3	0.421	6.563	2.636	5.563	1.318
11	1	0.424	7.680	3.294	10.120	1.735
11	2	0.391	10.040	5.872	10.400	3.051
11	3	0.347	14.640	8.447	23.480	8.511
12	1	0.466	7.963	5.161	11.259	2.548
12	2	0.446	9.840	7.097	11.480	2.226
12	3	0.347	13.040	8.000	17.800	4.723
13	1	0.233	18.074	11.135	41.000	10.112
13	2	0.223	20.037	15.723	42.778	16.191
13	3	0.205	24.593	18.371	50.037	24.019
15	1	0.500	7.480	4.440	7.160	1.720
15	2	0.500	9.885	5.346	10.077	1.654
15	3	0.369	7.833	4.780	10.042	2.220

Table 7 (*Continued*)

Therapeutic area ID	Period	Share OECD countries in network	Mean degree OECD countries	Mean degree non-OECD countries	Mean betweenness centrality OECD countries	Mean betweenness centrality non-OECD countries
16	1	0.511	6.304	4.136	6.522	1.409
16	2	0.545	9.208	5.150	7.417	1.850
16	3	0.444	8.208	3.767	11.292	2.767
17	1	0.403	9.407	7.900	14.481	6.200
17	2	0.419	10.769	6.361	13.192	1.944
17	3	0.325	14.480	10.135	19.640	5.000

Table 8
Network correlations period 2.

	Diff_DegreeCentrality	Collaboration	IncomeSimilarity	LanguageSimilarity	Diff_ResearchStrength	PointConnectivity	GeodesicCount
Diff_DegreeCentrality	1						
Collaboration	0.0994	1					
IncomeSimilarity	0.1053	0.1485	1				
LanguageSimilarity	0.2025	0.0779	0.1147	1			
Diff_ResearchStrength	0.6690	0.0386	−0.0019	0.1617	1		
PointConnectivity	0.3391	0.3479	0.3127	0.1513	0.4447	1	
GeodesicCount	0.2046	−0.0138	0.1560	0.0753	0.2799	0.3801	1

Table 9
Network correlations period 3.

	Diff_DegreeCentrality	Collaboration	IncomeSimilarity	LanguageSimilarity	Diff_ResearchStrength	PointConnectivity	GeodesicCount
Diff_DegreeCentrality	1						
Collaboration	0.0943	1					
IncomeSimilarity	0.0715	0.1507	1				
LanguageSimilarity	0.1719	0.0579	0.1217	1			
Diff_ResearchStrength	0.7047	0.0502	−0.0146	0.1306	1		
PointConnectivity	0.2731	0.3421	0.3123	0.0750	0.3938	1	
GeodesicCount	0.2149	−0.0152	0.1530	0.0526	0.2387	0.3140	1

References

Adams, J.D., Black, G.C., Clemmons, J.R., Stephan, P.E., 2005. Scientific teams and institutional collaborations: evidence from U.S. universities, 1981–1999. Research Policy 34 (3), 259–285.

Ahuja, G., 2000. Collaboration networks, structural holes, and innovation: a longitudinal study. Administrative Science Quarterly 45 (3), 425–455.

Ahuja, G., Soda, G., Zaheer, A., 2012. The genesis and dynamics of organizational networks? Organization Science 23 (2), 434–448.

Arora, A., 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., Schuster, H.G. (Eds.), Handbook of Graphs and Networks: From the Genome to the Internet. Wiley-VCH, Weinheim, pp. 69–84.

Barabási, A.-L., Albert, R., 1999. Emergence of scaling in random networks? Science 286 (5439), 509–512.

Bartholomew, S., 1997. National systems of biotechnology innovation: complex interdependence in the global system? Journal of International Business Studies 28 (2), 241–266.

Baum, J.A.C., Calabrese, T., Silverman, B.S., 2000. Don't go it alone: alliance network composition and startups' performance in Canadian biotechnology? Strategic Management Journal 21 (3), 267–294.

Borgatti, S.P., Halgin, D.S., 2011. On network theory? Organization Science 22 (5), 1168–1181.

Boschma, R., Frenken, K., 2010. The spatial evolution of innovation networks: a proximity perspective. In: Boschma, R., Martin, R. (Eds.), Handbook of Evolutionary Economic Geography. Edward Elgar Publishing, Cheltenham, pp. 120–135.

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

Bozeman, B., Corley, E., 2004. Scientists' collaboration strategies: implications for scientific and technical human capital? Research Policy 33 (4), 599–616.

Breschi, S., Catalini, C., 2010. Tracing the links between science and technology: an exploratory analysis of scientists' and inventors' networks? Research Policy 39 (1), 14–26.

Burt, R., 1992. Structural Holes: The Social Structure of Competition. Harvard University Press, Cambridge.

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

Calero, C., van Leeuwen, T.N., Tijssen, R.J.W., 2007. Research cooperation within the bio-pharmaceutical industry: network analyses of co-publications within and between firms? Scientometrics 71 (1), 87–99.

Cantner, U., Graf, H., 2006. The network of innovators in Jena: an application of social network analysis? Research Policy 35 (4), 463–480.

Coleman, J.S., 1988. Social capital in the creation of human capital. American Journal of Sociology 94 (Supplement), S95–S120.

Coleman, J.S., 1990. Foundations of Social Theory. The Belkap Press, Cambridge.

Cooke, P., 2006. Global bioregional networks: a new economic geography of bioscientific knowledge? European Planning Studies 14 (9), 1265–1285.

Cowan, R., Jonard, N., 2007. Structural holes, innovation and the distribution of ideas. Journal of Economic Interaction and Coordination 2 (2), 93–110.

de Solla Price, D., 1963. Little Science, Big Science. Columbia University Press, New York.

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

Dekker, D., Krackhardt, D., Snijders, T., 2007. Sensitivity of MRQAP tests to collinearity and autocorrelation conditions? Psychometrika 72 (4), 563–581.

Dosi, G., Pavitt, K., Soete, L., 1990. The Economics of Technical Change and International Trade. Harvester Wheatsheaf, New York.

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.

Furman, J.L., Kyle, M.K., Cockburn, I., Henderson, R.M., 2006. Public & private spillovers, location and the productivity of pharmaceutical research. Annales d'Économie et de Statistique 79/80, 1–24.

Furman, J.L., Porter, M.E., Stern, S., 2002. The determinants of national innovative capacity? Research Policy 31 (6), 899–933.

Gambardella, A., 1995. Science and Innovation: The US Pharmaceutical Industry During the 1980. Cambridge University Press, Cambridge.

Gassmann, O., von Zedtwitz, M., 1999. New concepts and trends in international R&D organization? Research Policy 28 (2–3), 231–250.

Glückler, J., 2010. The evolution of a strategic alliance network: exploring the case of stock photography. In: Boschma, R., Martin, R. (Eds.), Handbook of Evolutionary Economic Geography. Edward Elgar Publishing, Cheltenham, pp. 298–315.

Granovetter, M.S., 1973. The strength of weak ties? American Journal of Sociology 78 (6), 1360–1380.

Granstrand, O., 1999. Internationalization of corporate R&D: a study of Japanese and Swedish corporations? Research Policy 28 (2–3), 275–302.
Guellec, D., 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.
Hamilton, K.S., 2003. Subfield and Level Classification of Journals. CHI Research Inc., CHI No. 2012-R.
Henderson, R., Orsenigo, L., Pisano, G., 1999. The pharmaceutical industry and the revolution in molecular biology: interactions among scientific, institutional, and organizational change. In: Mowery, D.C., Nelson, R.R. (Eds.), Sources of Industrial Leadership: Studies of Seven Industries. Cambridge University Press, Cambridge, pp. 267–310.
Hoekman, J., Frenken, K., van Oort, F., 2009. The geography of collaborative knowledge production in Europe? The Annals of Regional Science 43 (3), 721–738.
Hubert, L.J., 1987. Assignment Methods in Combinatorial Data Analysis. Dekker, New York.
Hubert, L.J., Schultz, S., 1976. Quadratic assignment as a general data analysis strategy? British Journal of Mathematical and Statistical Psychology 29 (2), 190–241.
Jackson, D., 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., Barabási, A.-L., 2003. Measuring preferential attachment in evolving networks? Europhysics Letters 61 (4), 567–572.
Katz, J.S., Martin, B.R., 1997. What is research collaboration? Research Policy 26 (1), 1–18.
Kenis, P., Knoke, D., 2002. How organizational field networks shape interorganizational tie-formation rates? The Academy of Management Review 27 (2), 275–293.
Kim, J.W., Higgins, M.C., 2007. Where do alliances come from?: The Effects of upper echelons on alliance formation? Research Policy 36 (4), 499–514.
Knoke, D., Burt, R., 1983. Prominence. In: Burt, R., Minor, M. (Eds.), Applied Network Analysis: A Methodological Introduction. Sage, Beverly Hills, pp. 195–222.
Kogut, B., Shan, W., Walker, G., 1994. Knowledge in the network and the network as knowledge. In: Grabher, G. (Ed.), The Embedded Firm: on the Socioeconomics of Industrial Networks. Routledge, London, pp. 67–94.
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.
Lim, K., 2004. The relationship between research and innovation in the semiconductor and pharmaceutical industries (1981–1997)? Research Policy 33 (2), 287–321.
Lundberg, J., Tomson, G., Lundkvist, I., Skar, J., Brommels, M., 2006. Collaboration uncovered: exploring the adequacy of measuring university–industry collaboration through co-authorship and funding? Scientometrics 69 (3), 575–589.
Luukkonen, T., Persson, O., Sivertsen, G., 1992. Understanding patterns of international scientific collaboration. Science, Technology & Human Values 17 (1), 101–126.
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., Sundberg, C.-J., 2008. Intra-EU vs. Extra-EU scientific co-publication patterns in EU. Scientometrics 75 (3), 555–574.
Mayer, T., Zignago, S., 2006. Notes on CEPII's Distances Measures. CEPII (Centre d'Études Prospectives et d'Information Internationales), Paris.
McKelvey, M., Orsenigo, L., Pammolli, F., 2004. Pharmaceuticals analyzed through the lens of a sectoral innovation system. In: Malerba, F. (Ed.), Sectoral Systems of Innovation: Concepts, Issues and Analyses of Six Major Sectors in Europe. Cambridge University Press, pp. 73–120.
McPherson, J.M., 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., 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., Metcalfe, J., 2007. Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge? Research Policy 36 (5), 789–806.
Miquel, J., Okubo, Y., 1994. Structure of international collaboration in science-part II: comparisons of profiles in countries using a link indicator? Scientometrics 29 (2), 271–297.
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., 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., 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., 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., Brantley, P., 1992. Competitive cooperation in biotechnology: learning through networks? In: Eccles, R., Nohria, N. (Eds.), Networks and Organizations. Harvard Business School Press, Boston, pp. 366–394.
Powell, W., Grodal, S., 2005. Networks of innovators. In: Fagerberg, J., Mowery, D., Nelson, R. (Eds.), The Oxford Handbook of Innovation. Oxford University Press, Oxford, pp. 56–85.
Powell, W., Koput, K., 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., Owen-Smith, J., 1999. Network position and firm performance: organizational returns to collaboration in the biotechnology industry. In: Andrews, S., Knoke, D. (Eds.), Research in the Sociology of Organizations. JAI Press, Greenwich, pp. 129–159.
Powell, W., White, D., Koput, K., Owen-Smith, J., 2005. Network dynamics and field evolution: the growth of interorganizational collaboration in the life sciences? American Journal of Sociology 110 (4), 1132–1205.
Rivera, M.T., Soderstrom, S.B., Uzzi, B., 2010. Dynamics of dyads in social networks: assortative, relational, and proximity mechanisms. Annual Review of Sociology 36 (1), 91–115.
Rogers, E.M., 1995. Diffusion of Innovations. Free Press, New York.
Roijakkers, N., 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., 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.
Senker, J., Sharp, M., 1997. Organizational learning in cooperative alliances: some case studies in biotechnology? Technology Analysis & Strategic Management 9 (1), 35–52.
Shan, W., Hamilton, W., 1991. Country-specific advantage and international cooperation? Strategic Management Journal 12 (6), 419–432.
Sharp, M., Senker, J., 1999. European biotechnology: learning and catching-up. In: Gambardella, A., Malerba, F. (Eds.), The Organization of Economic Innovation in Europe. Cambridge University Press, Cambridge, pp. 269–302.
Ter Wal, A.L., 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., Vonortas, N.S. (Eds.), Innovation Networks in Industries. Edward Elgar Publishing, Cheltenham, pp. 81–104.
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., 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., 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.
Walker, G., Kogut, B., Shan, W., 1997. Social capital, structural holes and the formation of an industry network. Organization Science 8 (2), 109–125.
Wasserman, S., Faust, K., 1994. Social Network Analysis: Methods and Applications. Cambridge University Press, Cambridge.
Watts, D.J., Strogatz, S.H., 1998. Collective dynamics of 'small-world' networks? Nature 393 (6648), 440–442.
Witt, U., 2008. Evolutionary economics. In: Durlauf, S.N., Blume, L.E. (Eds.), The New Palgrave Dictionary of Economics. Palgrave Macmillan, Basingstoke.
Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge? Science 316 (5827), 1036–1039.
Zaheer, A., Soda, G., 2009. Network evolution: the origins of structural holes? Administrative Science Quarterly 54 (1), 1–31.
Zitt, M., Bassecoulard, E., Okubo, Y., 2000. Shadows of the past in international cooperation: collaboration profiles of the top five producers of science. Scientometrics 47, 627L 657.