Knowledge Structures and Complementarities in the Pharmaceutical Industry *

Moritz Müller[1], Robin Cowan[1,2], Geert Duysters[3,4], and Nicolas Jonard[5]

[1]Université de Strasbourg, France
[2]Universiteit Maastricht, The Netherlands
[3]UNU-Merit, The Netherlands
[4]Technische Universiteit Eindhoven, The Netherlands
[5]Université du Luxembourg

October, 2009

Abstract

This paper investigates how technological distance between firms affects their network of R&D alliances. Our theoretic model assumes that the benefit of an alliance between two firms is given by their technological distance. This benefit-distance relationship determines the ego-network of each firm as well as the overall network structure. Empirical relevance is confirmed for the bio-pharmaceutical industry. Although we find that the network structure is largely explained by firm size, technological distance determines the positioning of firms in the network.

Keywords: technological distance, research alliance, network formation, pharmaceutical industry

1 Introduction

Joint research and development (R&D) by two or more firms is frequent when firms face high innovation pressure and technological knowledge is dispersed among firms (Powell et al., 1996). Then the R&D alliance is important to generate technical innovations, because it governs the process of recombination of existing knowledge residing in different firms. [1]

*We are indebted to the ANR project AnCoRA for financial support of this work.
[1]See (McGee, 1995) for an historical account of technological novelty by recombination.

For a beneficial alliance it is crucial that both partners are able to evaluate each others knowledge and to appropriate the results of the alliance. Cohen and Levinthal (1990) termed this ability the absorptive capacity of a firm. Evaluation of foreign knowledge and integration into the own knowledge base is simpler when it is related to ones own prior knowledge. Therefore, due to increasing absorptive capacity, the benefit of joint R&D increases with knowledge relatedness.

However, the knowledge bases of the partners should not be too similar. Joint knowledge creation is valuable exactly when partners contribute knowledge new to each other and combine it in a new way. In principle, the opportunity to form novel combinations is higher the more diverse the knowledge bases of the firms are (Nooteboom et al., 2007). Hence, a higher cognitive distance between two firms yields a novelty gain.

The discussion shows the trade off between absorptive capacity and novelty gain. For a beneficial alliance, absorptive capacity and novelty gain are both preferred to be high. However, with increasing cognitive distance absorptive capacity decreases and novelty gain increases. This implies that benefit is maximized at some medium cognitive distance, the point of optimal cognitive distance (Nooteboom et al., 2007).

The concept of cognitive distance is very broad in that it incorporates any difference between the mind sets of the firms. Cognition includes not only the knowledge of facts but also e.g. interpretation, categorization and emotions. For R&D alliances technological knowledge seems to be most relevant and we may reduce the concept of cognitive distance to technological distance without losing too much insight. The implication remains the same: with increasing technological distance benefits of joint innovation first increase and then decrease. This has been tested empirically by (Mowery et al., 1998; Nooteboom et al., 2007). They found that joint R&D is most likely for pairs of firms having intermediate technological distances. [2]

Observing that the fundamental building block of a network is the bilateral alliance, the previous results invite the question whether the technological distance effect is visible in aggregate network structures. Could it determine the structure of the network and the position of firms therein?

The structure of an R&D network is likely to influence the generation and diffusion of knowledge in an industry. Specifically, (Cowan and Jonard, 2004) argue that small world networks (i.e. highly clustered networks with small path length (Watts and Strogatz, 1998)) foster knowledge accumulation of an industry. On the firm-level, empirical work shows that a firm's network position

[2] Similar, (Stuart, 1998) found that firms with higher technological overlap, measured by the share of common patent citations, are more likely to engage in strategic technology alliances. His study is particularly close to ours in that he explicitly considers the position of firms in technological space to infer on their number of alliances.

affects its knowledge sourcing and production behavior (Ahuja, 2000; Baum et al., 2000; Cockburn and Henderson, 1998; Gilsing et al., 2008; Powell et al., 1996; Shan et al., 1994). For example, a central position in the network gives a firm fast access to knowledge (Singh, 2005).

The question of how the distance-benefit effect between firm pairs contributes to the network structure has not been treated yet in the prior literature. In this paper, we investigate the question theoretically and empirically.

We model profit-maximizing firms forming alliances. Profits are determined by the distance-benefit relationship. Whereas the distance-benefit relationship is common to all firm-pairs, each firm pair has a specific technological distance. With the relationship and all distances given, we know the alliance decision of all firm pairs and the network is determined completely. Thus, the network characteristics for the individual firm and the overall network can be derived. The intuition gained from the model is that the position of firms in the knowledge space in combination with the benefit-distance relationship affects the network structure and the position of firms therein.

The model follows the connections model of (Jackson and Wolinsky, 1996) and its extension, the spatial social network of (Gilles and Johnson, 2000). Our model set up can be seen as a specification of the latter in that it models the benefit distance relationship to be inverse-U-shaped. However, in contrast to the literature on connections models we do not focus on stability and efficiency (Jackson et al., 2003) but rather on the network characteristics implied by the model.

The empirical relevance of our model is confirmed for R&D alliances in the bio-pharmaceutical industry. The analysis proceeds on three levels: i) on the dyad-level, technological distance is measured with patent data and the benefit-distance relationship is estimated. These estimates yield expectations of ii) the individual firm positions and iii) the network structure. We find that both, firm network positions and the global network structure, are affected by technological distance. However, the effect on the network structure is weak once the size of the firms is taken into account.

Thus, the paper finds that the benefit-distance relationship is a local effect of dyad formation which influences higher level network structures, especially on the firm level. This adds an economically motivated local effect to the toolbox of network analysis, which has hitherto been dominated by socially motivated local effects like referrals, trust or status (Powell et al., 2005).

The paper is organized as follows: section 2 presents the theoretic model from which the hypothesis are derived. Section 3 provides insight into the data, shows how technological distance is measured and the hypothesis are tested. Empirical results are given in section 4. The last section concludes.

2 Model

2.1 Technological Distance and Benefit

Consider a population of firms located in a knowledge space with a well-defined distance metric, t. Value resides not in firms but in alliances between firm pairs. Assume that the benefit of an alliance depends on technological distance in knowledge space but cost is fixed. Then, for two firms i and j having distance t_{ij} in the knowledge space, forming an alliance yields a benefit $f(t_{ij})$ and costs c. The alliance is valuable and hence formed, if $f(t_{ij}) > c$.

This set up can be seen as a simplification of the state space model (Gilles and Johnson, 2000) in two respects. Firstly, assuming that the value resides in the link instead of the agent takes away the necessity of specifying bargaining, side-payments and/or an allocation rule. It suffices to assume that both firms profit to some extent from a profitable link. Secondly, excluding that third party agents benefit from spillovers via indirect links allows for analytical derivation of networks with arbitrarily many agents. Otherwise, the benefit of forming a link would depend on previously existing links. Then, the order of decision making becomes relevant and each permutation needs to be considered, which limits the analysis to a small set of agents.

Both simplifications seem reasonable in the context of R&D alliances. The value can be considered as residing in the alliance, if the benefit stems mainly from the new knowledge generated or direct knowledge spillovers from the partner. Ignoring indirect spillovers in the model of link formation does not neglect their existence. Rather, it implies that firms decide upon the direct effects of the alliance. The subsequent analysis profits from these assumptions, because it helps to focus on the effect of optimal technological distance.

The discussion of the technological distance effect implies that the value of an alliance is an inverse-U-shaped function of distance. In mathematical terms, $f(t)$ is defined to be a continuous, differentiable, real-valued, single-peaked function, with t being the technological distance between two firms. Assume further that there exists a finite t^* such that $\forall t \geq t^*$, $f(t) = 0$; and possibly there exists a t^{**} such that $\forall 0 \geq t \geq t^{**}$, $f(t) = 0$. Because the value function is single peaked and costs are assumed to be constant, all alliances in some range $[a, a+b]$ are profitable and hence realized. Definitions of a and b follow from: $f(a+b) = c$ and $a = 0$ if $f(0) \geq c$, otherwise, a is defined as $f(a) = c$ (see figure 1).

By forming alliances the firms construct an alliance network. The network can be described as a graph g, in which the firms are nodes and the alliances are the links connecting the nodes. Different assumptions about the nature of the knowledge space and the distribution of firms therein lead to different networks. This paper examines a one-dimensional knowledge space, in which

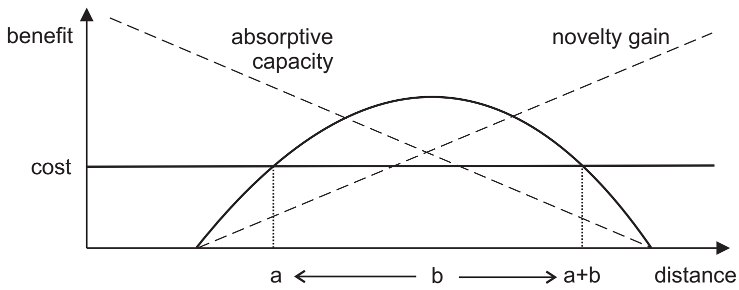

Figure 1: The inverse-U-shaped benefit-distance relationship arises from the trade-off between absorptive capacity and novelty gain (the figure displays a multiplicative effect). Taking into account the costs of alliance formation, one finds the range $[a, a+b]$ in which alliances are profitable. (Adapted from (Nooteboom et al., 2007))

firms are uniformly distributed. This simplifies the analysis but the intuition gained from the model can be easily extended to multi-dimensional knowledge spaces with unevenly distributed firms. We treat two cases: in the first case, the technological distance effect is illustrated on an infinite knowledge space. In the second case, boundaries restrict the knowledge space to be finite and thereby alter the technological distance effect.

In the first case, assume that the knowledge space is unbounded on the real line over which agents are uniformly distributed. In this case, the knowledge space is translation invariant, so agent 0, located at the origin, is a representative agent. This agent will maintain a link to agent i if and only if $f(i) \geq c$. Since $f(\cdot)$ is single-peaked this implies that there are values a and b such that agent 0 will form links with all agents $i \in [a, a+b] \bigcup [-a, -a-b]$ (see figure 1).

In the unbounded knowledge space, all agents face the same problem. Now suppose the knowledge space is bounded between 0 and 1. Then agents in the center are in a different position than those at the boundaries because the boundaries restrict the set of potential partners.

How the boundaries restrict the neighborhood of firms can be seen in figure 2, e.g. the upper left graph. Consider links to the right of the agent. For the agent at $i = 0$, its neighborhood will run from a to $a + b$. As we increase i, the neighborhood remains unrestricted until $i + a + b > 1$, or equivalently, $i = 1 - a - b$. As we increase i further, the right boundary restricts the neighborhood of agent i to be $[i + a, 1]$. Finally, at the point $i = 1 - a$, agent i no longer has any neighbors to the right. The partnering problem is symmetric to left and right, the same effect moving from $i = 1$ to $i = 0$ is seen for left side neighbors. This effect drives all the results on network measures in the bounded technological space in the next section 2.2.

Because the assumptions determine the network completely, in principle any network characteristic can be derived for individual firms as well as the whole network. In this paper however, we will focus on degree centrality, closeness centrality and clustering; three of the most common measures used in network analysis. The following derivations use intuitive arguments; they are based on mathematically rigorous demonstrations given in the appendix.

2.2 Network Measures

2.2.1 Degree Centrality

Degree Centrality of a node is the number of links it has to other nodes in the network. A firm with many R&D alliances is highly engaged in knowledge generation (Ahuja, 2000). From a resource based view, the alliances signal access to the knowledge or other resources residing in the partnering firms (Arora and Gambardella, 1990). The network degree distribution is commonly used to show the centralization in the network (Wasserman and Faust, 1994).

In the unbounded knowledge space all firms are in the same situation. The agent at the origin, 0, forms links with all partners $j \in [a, a+b] \cup [-a, -a-b]$. Assuming firms are uniformly distributed with density one, the size of the neighborhood of agent 0 is 2b. Because all firms are in the same situation, the degree distribution of the graph is a point mass at $2b$.

In the bounded knowledge space the degree of firm i depends on its position in combination with the benefit range $[a, a+b]$. If $a + b < 1$, some firms near the left boundary are not restricted on the right and will have a full right neighborhood of size b. When moving to the right, first agents are restricted in their right neighborhood and finally the boundary at one prevents completely a right neighborhood.[3]

As the right boundary becomes more restrictive, the left boundary lessens. Eventually agent i realizes a left neighborhood when moving from position 0 to the right. Whether the gain of lefthand neighbors is higher than the loss of righthand neighbors depends on the size of the minimum and the maximum distance, i.e. a and $a + b$.

Figure 2 shows both cases: if i) $a + (a + b) > 1$, agents moving away from zero restrict their right neighborhood before a left neighborhood forms. In this case, being more central in knowledge space implies lower degree centrality. If

[3] Assuming that firms are uniformly distributed with density one over the $[0, 1]$ interval in fact is not sensible, because it implies that only one firm is in the knowledge space. Nevertheless, the results can be applied to an arbitrary number of agents simply by scaling the density.

ii) $a + (a + b) < 1$, agents moving away from zero form a left neighborhood before their right neighborhood becomes restricted. In this case, agents which are central in knowledge space are also central in their degree.

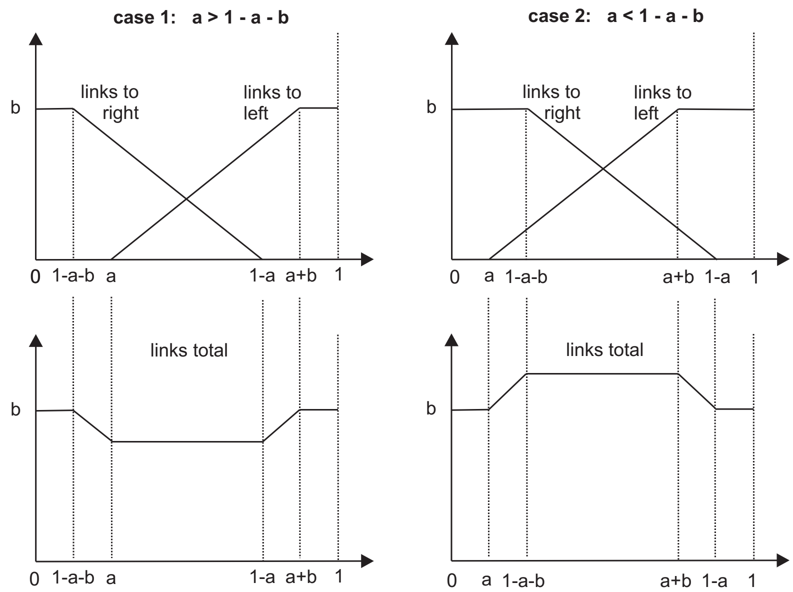

Figure 2: In the bounded knowledge space (here between 0 and 1) the number of links of the agent depends on its position. Firms with a central position in the knowledge space are less (more) central in the network in case 1 (case 2).

When the degree of each node is known, the degree distribution is gained simply by sorting the nodes according to their degree. Although the two cases ($a > 1-a-b$ and $a > 1-a-b$) imply different levels of link formation (the first case being lower), they do not imply a qualitative difference in the shape of the distribution. In both regimes, we are going to find a skewed degree distribution where some agents have many and some agents have few links.

2.2.2 Closeness Centrality

Closeness Centrality of a node is the average distance to all other reachable nodes:

$$closeness_i = 1/\left(\frac{1}{N}\sum_{j=1,j\neq i}^{N} d_{ij}\right)$$

where d_{ij} is the shortest path (i.e. the minimum number of links) connecting two vertices i and j in the network. This measure is critical when links guide the information flow in the network and the information content is decreasing with distance. In that situation, firms in a central position have good access to information and might be influential emitters of information (Singh, 2005). High average closeness in a network indicates fast spread of information.

This measure is not reasonable when the knowledge space is unbounded, because each agent will have infinitely long paths. When knowledge space is bounded, however, distances are all finite and the two cases displayed in figure 2 are qualitatively different.

First, consider case 1 where the minimum distance and the benefit range are large ($a + (a + b) > 1$). Agents on the left with $i \in [0, 1 - a - b]$ are completely restricted to the left but not to the right. Thus, they form direct links to right-hand agents $j \in [i + a, i + a + b]$, which means a shortest path of 1. All other agents ($j \in [0, i]$, $j \in [i, i+a]$ and $j \in [i+a+b, 1]$) are reached in two steps (via the agents $i + a$, $i + 2a$ and $i + b$ respectively). Thus, the average path length is $1\ b + 2\ (1 - b) = 2 - b$ and closeness is $1/(2 - b)$.

As i increases towards $i = 0.5$, agents $i \in [1 - a - b, a]$ increase their average path length. Because the neighborhood to the right becomes more and more restricted, the mass of directly connected agents decreases and the mass of agents reached via two links increases. In total, the average path is $1 + i + a$, which increases with i. Closeness centrality, the inverse of average path length, thus decreases as firms move away from the boundary. At position a, we reach the floor with $1/(1 + 2a)$. From this point a left neighborhood forms and the effects of the increasing left neighborhood and the decreasing right neighborhood cancel.

As was the case with degree centrality, we get the opposite result in case 2, where minimum distance and the benefit range are small ($a < 1 - a - b$). Agents near the boundary, $i \in [0, a]$, are in the same situation as agents near the boundary in case 1 and have an average path of $2 - b$. However, because a is relatively small, firms $i \in [a, 1 - a - b]$ connect to left hand agents directly before the righthand neighborhood is restricted. This reduces their average path length to $(2 - b) - (i - a)$. From position $i = 1 - a - b$ the average path length stabilizes to $1 + 2a$. Thus, in the second case closeness centrality is highest when firms are more central in knowledge space.

The results for closeness centrality parallel the findings for degree centrality: firms being more central in knowledge space have lower or higher closeness centrality, depending on the minimum (a) and the maximum distance ($a + b$). Both cases imply that closeness distribution is skewed.

2.2.3 Clustering Coefficient

The clustering coefficient of a node quantifies how close the immediate neighborhood of a node is to being fully connected. A firm partnering with firms which are otherwise unconnected has a low clustering coefficient. Burt [cite missing] argues that such a position is to be preferred because this firm controls the information flow and has potentially a strong bargaining position. On the other hand, in a small-world network average clustering and average closeness both is high (Watts and Strogatz, 1998). Whereas high clustering might foster knowledge generation due to specialization of groups of firms, high closeness enables fast diffusion of knowledge (Cowan and Jonard, 2004).

The clustering coefficient of a node is defined as the number of links among its neighbors divided by all links that possibly could exist among them:

$$clustering_i = \frac{2|\{e_{jk}\}|}{|N_i|(|N_i|-1)} \ ,$$

where the neighborhood N_i is the set of neighbors of i, i.e. those agents, i is directly connected with, and $\{e_{jk}\}$ is the number of realized links among the neighbors. When the two neighbors, j and k, link together, they close a triad with agent i. Therefore, $\{e_{jk}\}$ gives the number of triads agent i is involved in. This is a network measure in its own right, which may be used to indicate the effect of referrals in a network.

In the unbounded knowledge space one might again consider agent 0 as the representative agent. Agent 0 has a neighborhood of size $2b$, equally divided into a left and a right neighborhood of size b. A fully connected graph of size $2b$ contains $1/2(2b)^2 = 2b^2$ links. However, the minimum distance a prevents the agents in the right (left) neighborhood to fully connect among each other. Therefore, the right (left) neighborhood yields only $1/2(b-a)^2$ links. Similarly, the benefit range prevents some connections among left and right neighbors. For example, the right-most left agent $-a$ reaches only right neighbors in the range $[a, b]$. Therefore, the number of links between left and right neighbors is $1/2(b-a)^2$.

Thus, the number of links among all neighbors is $3/2(b-a)^2$. Dividing by $2b^2$, the number of links in a complete graph formed by a neighborhood of size $2b$, yields a clustering coefficient of $\frac{3}{4}\frac{(b-a)^2}{b^2}$. In the unbounded knowledge space this is the same for all agents.

In the bounded knowledge space, the boundaries may restrict the left and/or the right neighborhood of an agent. Again, this makes the clustering coefficient a function of the position in knowledge space. However, in contrast to degree and closeness, here we find that firms which are more central in knowledge space will always have a lower clustering coefficient than firms at the boundary.

First consider case 1 with $a > 1 - a - b$. Firms at the boundary only have a right neighborhood, of size b, and thus a clustering coefficient of $(b-a)^2/b^2$. In the range $[1-a-b, a]$, firms become increasingly restricted in their right neighborhood. This implies that both the number of realized links and the number of potential links among the neighbors decrease. However, the number of realized links decreases faster, because the minimum distance a which prevents that all potential links are realized becomes more important in a smaller neighborhood. From agent a on, the total size of the neighborhood remains the same when moving towards the center. However, the right and left neighborhood become more symmetric. Because the number of realized links is related to the square of the right and the square of the left neighborhood size, higher symmetry never increases the number of realized links. Thus, the clustering coefficient further decreases and, depending on the specification of the benefit range, eventually remains stable.

The same effect can be observed in case 2 with $a < 1 - a - b$. Again, firms start with a clustering coefficient of $(b-a)^2/b^2$ at the boundary. In the range $[a, 1-a-b]$, both the number of realized links and the number of potential links among the neighbors increase. However, because the symmetry of the left to the right neighborhood increases along with the total neighborhood, the clustering coefficient decreases. Moving from $1 - a - b$ to the center, left and right neighborhood become equally sized whereas the total neighborhood remains stable. Then, depending on the exact specification of the benefit range, the clustering coefficient remains stable or to decreases further.

Unlike degree and closeness, the clustering coefficient is always lower for firms in the center of the knowledge space. This is due to the normalization by the potential number of links in the neighborhood. However, the two cases make a difference for the number of triads a firm is involved in: firms in the center have a lower (higher) number of triads in the first (second) case. Therefore, the number of triads is considered next to the clustering coefficient in the empirical section.

2.3 Hypothesis

The implication of the model is that when the distance benefit range is small (large) relative to the technological space, a firm which is central in technological space is more (less) central in the research network. An obvious way to proceed would be to determine the relevant case for a population of firms and test the implication of the model directly. To this end, one need to measure the distance benefit range, the diameter of the knowledge space and how central a firm is in knowledge space.

However, whereas in the model firms are uniformly distributed in a one-dimensional space, in reality we are confronted with unevenly distributed firms in a multi-dimensional space. But what is the diameter of a multi-dimensional rectangular? And how should distance and firms be weighted to calculate the center? Answers to these questions seem arbitrary.

Therefore, we do not test directly the relationship between central positions in technological space and firm network positions as well as global network structures. Although the empirical section provides statistical evidence on this specific implication, the hypotheses rather follow the analysis of the model. In detail, we derive three hypothesis: the first hypothesis tests the key-assumption of the model, which is the distance-benefit relationship on the dyad level. The second hypothesis tests its implications on the network characteristics on the firm level. Finally, the third hypothesis tests the implications on network distributions on the global level. This makes hypothesis testing independent from the dimension of the knowledge space and furthermore allows for disentangling the assumption on the benefit-distance range from its effect on higher level network structures.

The main assumption of the theoretic model is that the alliance formation of any two firms depends on their technological distance. Because the literature suggests specifically an inverse-U-shaped function, we formulate:

Hypothesis 1 *The probability of two firms forming an alliance will be a curvilinear function of their technological distance.*

Note that the functional form itself is not crucial to make the model work. As long as there is a benefit-distance relationship, one might consider it as a local network effect. The model shows how such a local network effect would determine the network if it was the only effect at work. For three network measures, degree, closeness and clustering, the resulting network has been analyzed on two levels: on firm level and network level.

On firm level, network characteristics of individual firms became a function of their position in knowledge space. Depending on the exact specification of the benefit-distance relationship, higher centrality in knowledge space implied higher or lower centrality in the network as well as lower clustering. Thus, knowledge of the firms' position in network space and of the benefit distance relationship should enable us to infer at least tendencies of individual firms' network characteristics:

Hypothesis 2 *Firm level network characteristics depend on the firms' position in knowledge space for a given benefit-distance relationship.*

On the network level, network measures describe the architecture of a network by neglecting the individuality of the nodes. The analytical derivations showed that depending on the specification of the local effect network distributions will be more or less skewed and on a higher or lower level. The relevance of the model for the network architecture is tested by

Hypothesis 3 *Distributions of network measures are related to the firms' distribution in knowledge space for a given benefit-distance relationship.*

The hypotheses are formulated in broad terms to capture the main idea of the model: the benefit-distance relationship is a local effect, which determines the alliance decision of firm pairs. Because the network is the aggregate of all alliance decisions, the local effect shapes firm network characteristics and network distributions.

3 Empirical Methods

3.1 The Pharmaceutical Industry

The hypotheses are tested on the pharmaceutical industry, because the theoretical model is expected to be especially relevant for this industry. Firstly, the alliance network is large and half of the alliances focus on joint research & development. Secondly, firms possess distinctive technological competences. Both can be traced back to the biotechnology revolution (Arora and Gambardella, 1990; Galambos and Sturchio, 1998; Henderson et al., 1999; Orsenigo et al., 2001). Starting in the late 1970s, the emergence of a wide array of new scientific disciplines in life sciences lead to various new methods and processes. Because the scientific advances originated outside the established firms in universities and public research organizations, the technological change induced industrial change. The population of firms changed because biotech start-ups entered the industry. They were typically founded by researchers to commercialize their scientific discoveries and therefore are based on specific technological competence. For the established pharmaceutical firms, one important pathway to adopt the new technological competences have been alliances with new specialized firms.

Nowadays all pharmaceutical firms are based on modern life sciences (Cockburn et al., 1999). Nevertheless, research alliances remain important in the industry. No firm is able to master all the fields which are potentially relevant for the development of new drugs. Therefore firms need to specialize and when necessary join complementary technological knowledge in research alliances (Powell et al., 2005). This makes the pharmaceutical industry a promising candidate for an empirical application.

3.2 The Sample

The firm sample is drawn from the CGCP database. The CGCP database is a comprehensive collection of publicly announced formal agreements. [include perhaps overall coverage time, industries, types]. A valuable feature is that it classifies alliances by industry and type (such as e.g. joint venture, commercial or research alliance).[4] The classification allows us to focus on research and de-

[4] For a description see www.cgcpmaps.com .

velopment alliances in the pharmaceutical industry.

The sample consists of the 250 firms, being most active in the pharmaceutical industry. To derive the sample, first all dyadic (bio-)pharmaceutical alliances between the years 2001 and 2006 (inclusive) have been extracted. Because firm level information needs to be added, not all firms involved could enter the sample. Selecting the 250 firms having most alliances assured to get a dense network with many alliances among the selected firms.

This sample is not representative; neither of the pharmaceutical industry nor of the global pharmaceutical network. However, the dependent variable is the alliance decision of the firm-dyad and not the number of alliances of the firm. Because selection is not based on the dependent variable, estimates need not be biased.

The technological position of firms is measured with patent data. The advantages and disadvantages of measuring technological capabilities with patent data have been discussed elsewhere (e.g. (Pavitt, 1982)). Because in the pharmaceutical industry firms patent extensively (Arundel and Kabla, 1998), we think that patent information reflects sufficiently the technological activity of the firms. The objectivity, information content and availability of patent data makes it superior to other information sources in our case.

The patent data has been extracted from the EPO Patstat database (EPO, 2008). Only those patents, which seem relevant for the bio-pharmaceutical industry have been considered. The restriction is based on concordances of the international patent classification (IPC) on four digit level to the biopharmaceutical industry. In detail, the set of IPC classes considered comprises those of the OECD definition (OECD, 2008b), the MERIT definition (Verspagen et al., 1994) and the ISI definition (Schmoch et al., 2003). One invention often is patented via a priority application to a national office and equivalent foreign versions of the application. In these cases, double counting has been avoided by considering only the priority application (OECD, 1994).

The hypotheses imply a direction of causality, namely that a firm's technological characteristics effects its alliance activity. This is accounted for by sampling the patent data from a time period previous to the time period of the alliance data. Whereas the alliances took place between the years 2001 and 2006, the patents have a priority date between the years 1995 and 2000.

The firm names given in the alliance database denote mostly a pharmaceutical business, either the entire group or a subsidiary. Therefore patents have been matched on the same level when possible. In those cases, where the pharmaceutical business is part of a diversified group but applies for patents solely in the name of the group no matching can be done. Additionally, for some firms no patents have been found due to the time or IPC restriction.

The patent matching yielded patent applications within the given priority date and IPC classes for 212 firms or their respective pharmaceutical business. For ten firms, mostly software and service firms, no patents could be found at all. Six firms only applied for patents on behalf of a diversified group. Twenty-two firms applied for patents but after the given year span.

In order to control for firm size, the number of employees has been collected from publicly available information, mostly annual reports of the SEC. About seventy per cent of the figures are at or before 2001. For the rest of firms this information could only be obtained from later years. For 14 out of the 250 firms, the number of employees could not be gained.

Thus, the final sample consists out of 250 firms. 38 firms have zero or missing patent assignments, 14 firms have missing employee information and 45 firms have neither patent nor employee information assigned. Finally, for 205 firms, which is 82% of the sample, patenting and employee information is given and these firms constitute the sample we work with.

3.3 Measures

3.3.1 Joint technological agreement

The dependent variable on the dyad-level, joint technological agreement (*joint-tech*), is defined as a joint project of two firms, in which both firms contribute to research and/or development. This definition excludes for example research projects conducted by one firm and financed by another. The fact that only publicly announced agreements enter the CGCP data base inevitably imposes a restriction to formal agreements.

3.3.2 Firm network position and network structure

The dependent variable on the firm-level is the firm network position. On the network-level it is the network structure. The firm network position as well as the network structure are described using the four network measures degree centrality, closeness centrality, clustering coefficient and the number of triads a firm is involved in. (Short notations are *degree*, *closeness*, *clustering* and *triads* respectively (for definitions see 2.2)). All network measures are calculated from the network of joint technological agreements among the firms, for which the distance-benefit relationship is estimated.

3.3.3 Technological position

The technological position of a firm is given by its technological distance to all other firms. The technological distance between any two firms is measured on their patent portfolios, where we take into account the size of the patent portfolios as well as the technological classes covered by the portfolio.

During examination, a patent examiner of the patent office assigns each patent according to the inventions claimed to one or several technological classes of the international patent classification (IPC) (OECD, 1994, page 30). Therefore, the IPC classes of a firm's patents reveal in which technological fields a firm is active. For indication, we use the main and secondary IPC classes. Naturally, two firms are technologically close when they patent in the same technological fields. The IPC overlap measures how close two firms are. It is the number of IPC classes covered jointly by both firms divided by the number of IPC classes covered by at least one firm:

$$overlap_{ij} = \frac{|IPC_i \bigcap IPC_j|}{|IPC_i \bigcup IPC_j|},$$

where IPC_i is the set of IPC, in which firm i had at least one patent applications and $||$ denotes the size of the set. In order to allow for a curvilinear relationship the square of the overlap ($overlap_{ij}^2$) is included in the estimations as well.

The overlap measure loses information on the size of the patent portfolios. Therefore, the complete information on the size of the patent portfolios of firms i and j is captured by two further variables: the sum and the absolute difference of the log-transformed patent count of firm i and j ($absDiffLnPC_{ij}$ and $sumLnPC_{ij}$).

Note that $absDiffLnPC_{ij}$ and $sumLnPC_{ij}$ are information equivalent to two variables indicating the log transformed patent count of the smaller and the bigger portfolio. The number of patents is log-scaled in order to take into account the decreasing importance of one more patent in a bigger patent portfolio. Technically, the log-scale leads to less skewed distributions.

In the literature also other distance measures based on patents have been used. (Mowery et al., 1998; Schoenmakers and Duysters, 2006) calculated the overlap of patent citations. The information on technological classes so far entered the cosine index (Jaffe, 1986, 1989), the correlated revealed technological advantage (cRTA) (Cantwell and Colombo, 2000; Gilsing et al., 2008; Nooteboom et al., 2007) or the euclidean distance (Rosenkopf and Almeida, 20030601). Some of these will be considered in the sensitivity analysis.

3.3.4 Firm size

Features of drug development and commercialization hint to further drivers of alliance formation (Galambos and Sturchio, 1998; OECD, 2008a). Development of new drugs is extremely costly and time consuming. In average 800 million dollars need to be expendet over 10 years to bring a new drug to the market.

Drug application processes are country-specific and demand strong organizational competencies to meet legal requirements. Because production costs are low compared to the high initial development expenses, sales revenues need to be maximized. This can only be achieved with strong marketing and distribution channels in the national markets.

Pharmaceutical firms are heterogeneous in their access to technological, financial and organizational resources (Pfeffer and Nowak, 1976; Eisenhardt and Schoonhoven, 1996). Therefore, research and development alliances are also motivated by financial and organizational interdependencies and these are especially strong between small and large pharmaceutical firms.

Because the size of the patent portfolio is strongly correlated with the size of the firm, controlling for firm size is crucial to sort out technological from financial and organizational interdependencies. This is achieved by introducing the two variables $absDiffLnEmployees_{ij}$ and $sumLnEmployees_{ij}$ combining the size information of two firms i and j. As for the size of the patent portfolios they denote the sum as well as the absolute difference of the log-transformed number of employees of two firms i and j.

3.4 Analysis

3.4.1 Statistical analysis

To test the first hypothesis, the local effect of technological distance on joint technological agreements is estimated. The estimates assign to each firm-pair a probability of forming an alliance and, when aggregated, yield expections on the network. The expected network implies expected firm network positions and expected network distributions. The relevance of the local effect on the network is revealed by comparing the expected with the observed firm network positions (hypothesis two) as well as the network distributions (hypothesis three). The next paragraphs discuss these steps in more detail.

A logit function is an appropriate model for the decision of two firms to form an alliance. However, when estimating link formation in a network, the non-independence of observations is a problem (van Duijn and Vermunt, 2006). An important source of dependence is the repeated observation of one firm over several firm-pairs. This is likely to cause correlation of estimated errors over firm-pairs, because some firms are more susceptible to form alliances then others for unknown reasons. Then, maintaining the independence assumption reduces the standard errors unduly and potentially gives biased coefficient estimates.

This problem is similar to that of repeated observations of one firm over time in a panel. In the panel setting, the problem is usually handled by introducing unobserved firm specific effects (Cameron and Trivedi, 2005). Under the assumption that firm specific effects are uncorrelated with other independent variables, one estimates a random effects model. When correlated, the random effects model yields biased coefficient estimates and the less efficient fixed effects model is appropriate. Which model to choose is decided upon a Hausman test, which tells whether the coefficients can be assumed to be equal given their variances.

We apply the standard solution for panel data to the estimation of link formation in a network. [5] Different from panel data we are handling dyads. Therefore each equation contains not one but two unobserved firm specific effects. As common for panel data models, we distinguish random and fixed effects.

The fixed effects model is estimated simply by introducing a dummy variable for each firm (Stuart, 1998). This does not cause the incidental parameter problem because the number of firm pairs (observations) increases much faster then the number of firms (variables). However, firms which have no links with other firms in the sample need to be excluded, because their fixed effect is minus infinity (not defined). This is not a problem for the random effects model, where the inclusion of these firms rather increases the variance of the random effects distribution. A random effects model has been proposed by (Hoff, 2003). We estimate it by maximum simulated likelihood under the assumption that firm specific effects are independent, normally distributed.

Because a Hausman test showed that the coefficient estimates of the random effects model are similar to those of the fixed effects model, we present only the results of the more efficient random effects model; estimated on the complete sample. Econometric details and results for fixed and random effects estimation are given in the appendix.

Introduction of firm specific effects does not necessarily make observations independent. Errors might still be systematically correlated, for example when firms favor alliances with firms being already close in the network. One strategy is to incorporate sufficient statistics for different kind of dependencies, as in the framework of Markov Graphs (van Duijn and Vermunt, 2006). The problem is that estimation might not be possible for some (larger) networks (Hunter et al., 2007), which happened in our case when introducing statistics of a dyadic dependence model. Because firm specific effects probably control for the most important source of bias and variance deflation, we leave the problem of more complicated network dependencies to future research.

The estimates obtained from the logit model are used to form expectations on the firm network positions. In principle expectations can be analytically

[5] This is advantageous to correcting the deflated standard error (Fafchamps and Gubert, 2007) of the logit estimation without specific effects, because it is more efficient and takes into account that estimates are possibly biased.

derived. For example the expected degree centrality of a firm is simply the sum of the probabilities of link formation over all firm dyads the firm is involved in. Analytical derivation of the expectations of the other network measures is more complicated but easily obtained by simulation. One instance of a network is simulated by random realization of all links given their probability to be formed. From each simulated network the position of each firm in terms of degree, clustering, closeness and number of triads is calculated. Then, the average over all simulations yields the expected firm position. The presented expectations are based on 1000 simulations, so that different simulation runs give the same results. Significant correlation of the expected network positions with the observed ones verifies hypothesis two.

Hypothesis three is similarly tested by comparing the expected with the observed network distributions. From each simulated network the network distributions are obtained. Their average gives the expected network distributions. Visual comparison of the distributions is valuable to judge the validity of hypothesis three (Hunter et al., 2008). In addition, we provide the Kullback Leibler Information Criterion (KLIC) (see e.g. (Cameron and Trivedi, 2005)).

3.4.2 Data analysis

In total there are 205 firms for which patent and size information is given. Crossing all firms yields 13695 firm-pairs which are used for estimation. These firm-pairs joined for 332 technological agreements, corresponding to 2% of all potential links.

Firms contributed unequally to link formation in the observed network. Whereas 39 firms have no links with other firms in the sample and 41 firms have one link only, five firms have fourteen or more links within the network. The high degree centralization is coupled with high closeness centralization. Clustering in the network is low compared to other industrial networks (checkref); we observe not more than 27 triads.

Figures 3 and 4 give a coherent picture of the case in which centrality in technological space leads to high centrality in the network and a higher number of triads. Figure 3 positions the firms in a (two-dimensional) knowledge space and displays their research network. It seems that firms being more central in knowledge space are also more central in the network. Technological distance of most alliances is rather short. Most alliances span only half the technological space and there are no alliances which span the entire space. This means that the benefit-range is small relative to technological space, which implies the second case derived in section 2.2.

This is confirmed in figure 4, where the average overlap of one firm to all other firms measures the firm's position in technological space. Firms with large

Figure 3: The network of joint technological agreements. Nodes are mapped into two dimensional knowledge space based on firm-pair overlap using the fruchterman reingold algorithm. Size of the nodes equals the log-transformed number of agreements within the network.

average overlap are close to most other firms and, therefore, can be considered to be close to the center of the technological space. The top left panel in figure 4 plots degree centrality as a function of average overlap. It seems that the population is divided at an average overlap of 0.3. There are only two firms which are distant from the technological center (average overlap below 0.3) and yet have more than five alliances in the sample. Although there are firms which are close to the center in technological space (overlap above 0.3) and have few alliances, firms closer to the technological center in general have more alliances. The absence of any firm being at the boundary of technological space and having a high degree even suggests that being central in technological space is a prerequisite for having many research alliances.

The effect on closeness centrality, shown in the top right panel in figure 4, is less clear cut. Here, the Pearson's correlation, with a coefficient of 0.34 and a significance level below 0.001%, gives a clear indication. Again, firms located near the center of the technological space tend to have high closeness centrality.

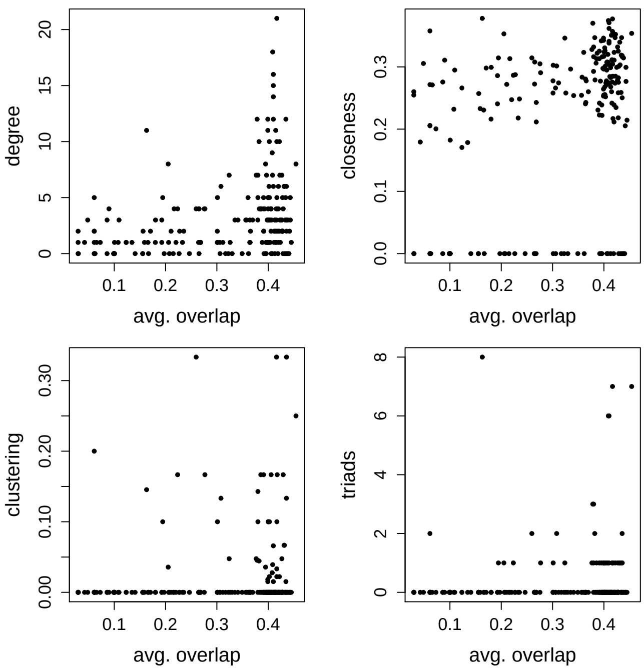

Figure 4: Average overlap versus firm network characteristics. The higher the overlap to all other firms in the sample, the more central a firm is in technological space. Firms having a degree or closeness centrality of zero are singletons, not connected to the network.

However, the clustering coefficient, given in the bottom left panel, deviates from theoretical prediction in that it is higher for firms being close to the technological center. The reason can be found in the bottom right panel, which displays the number of triads. Whereas in the model firms at the boundary of technological space have a higher clustering coefficient due to having fewer triads and even lower degree, in our sample they are involved in practically no triads and, therefore, their clustering coefficient becomes zero.

Thus, our statistical evidence is internally consistent with the model results. Specifically, the observed relationship of the firms' position in technological space and their network characteristics is implied by the second case of the model. Hypothesis testing, in the next section, is oriented on the principal

		mean	std.dev	1	2	3	4	5	6
jointtech	(1)	0.02	0.13						
sumLnPC	(2)	8.77	2.94	0.09***					
absDiffLnPC	(3)	2.38	1.75	0.06***	0.14***				
overlap	(4)	0.32	0.23	0.04***	0.57***	-0.37***			
overlapSq	(5)	0.15	0.17	0.04***	0.51***	-0.35***	0.95***		
sumLnEmpl	(6)	13.11	3.41	0.07***	0.56***	0.23***	0.22***	0.23***	
absDiffLnEmpl	(7)	2.74	2.05	0.08***	0.16***	0.34***	-0.15***	-0.18***	0.29***

Table 1: 20910 firm pairs from crossing 205 firms. Mean, standard deviations and pearson correlations for variables used in estimations. *,**,*** signify 5%, 1% and 0.1% rejection levels of significance.

steps of the model analysis and takes into account additional variables. Therefore, the independent variables need to be transformed and then combined on dyad level.

The distribution of number of patents is extremely right skewed, as is the number of employees. Number of patents range from one patent for eleven firms to 10500 patents for one firms, with a median of 62 and a mean of 591. The histogram becomes symmetric in log-scale with median being 4.2 and mean 4.3. Sizes of the firms ranges from 5 to 120000. Again, log-transformation centers the histogram around a value of 6.

The patent and employee information is used to construct the independent variables describing the dyad. Overlap is slightly right skewed with 11.1% of dyads having no overlap and 0.2% having complete overlap. Since overlap is mostly between 0 and 1, it is a valid metric which is capable of differenciating the firm-pair distances. Because the number of patents and employees have been log-scaled before being summed and differenced, the resulting variables all have a smooth distribution ranging between 0 and 30.

Table 1 shows that all variables are significantly correlated. The high significance is partly the effect of inflating the observations by forming firm-dyads. Nevertheless, all technological indicators are highly correlated with jointtech, supporting the importance of technological characteristics for joint technological agreements. However, the correlation of the employee information with technological characteristics hint to organizational and financial drivers of alliance formation and the importance to control for such drivers.

4 Results

4.1 Hypothesis Testing

4.1.1 Hypothesis 1

Table 2 reports the results of the regression analysis on the local effect of technological distance on joint technological agreements. The estimations support the first hypothesis. There is a curvilinear relationship between our structural measure of technological distance, *overlap*, and joint technological agreements, *jointtech*. Furthermore, we find a preference to combine with unequal partners regarding the size of the patent portfolio as well as firm size.

In table 2, model 1 is the baseline equation, containing the firm size control variables. The sum and absolute difference of log-employees (*sumLnEmpl* and *absDiffLnEmpl*) are positive, showing that big and small firms are likely to ally. This supports previous findings on the interdependencies of small and big firms in the pharmaceutical industry (Powell et al., 2005). Model 2 adds the sum and absolute difference of patent portfolio sizes (*sumLnPC* and *absDiffLnPC*). Their significance and a decreasing Akaike Information Criterion (AIC) assigns high relevance to both variables. The decrease of the size control variables supports the idea that the interdependencies between big and small firms are partly technological. Model 3 supports hypothesis one of a curvilinear relationship, with *overlap* being positive and *overlap*^{2} negative.

	model 1	model 2	model 3
intercept	-7.28*** (0.349)	-7.96*** (0.372)	-8.46*** (0.421)
overlap	–	–	4.05*** (0.989)
overlap^2	–	–	-2.63** (1.039)
absDiffLnPC	–	0.09** (0.034)	0.21*** (0.042)
sumLnPC	–	0.2*** (0.026)	0.1*** (0.032)
absDiffLnEmpl	0.25*** (0.026)	0.21*** (0.029)	0.22*** (0.031)
sumLnEmpl	0.16*** (0.019)	0.05* (0.024)	0.07** (0.024)
$\sigma^{2(1)}$	0.3 (0.097)	0.4 (0.101)	0.32 (0.1)
converged	converged	converged	converged
AIC	3218.09	3146.58	3124.94

Table 2: Random effects logit models with dependent variable *jointtech*. 20910 firm pair observations from crossing 205 firms. Standard errors in brackets; *,**,*** signify 5%, 1% and 0.1% rejection levels of significance. (1) The estimate of random effects variance follows a log-normal distribution and are therefore strictly positive.

In order to test hypothesis 1, three models estimated a local effect of network formation: the first model the heterophily of big and small firms, the second model adds the heterophily of firms with big and small patent stocks and the third model adds the distance benefit relationship of technological distance. All local effects are significant - separately and jointly.

4.1.2 Hypothesis 2

Hypothesis 2 proposes that network characteristics of a firm depend on its position in the knowledge space and that the relationship is determined by the benefit-range. Based on simulation of networks using the model estimates gained above, we derived the expected network position of each firm. Correlation of expected with observed network positions shows how well the respective model of dyad formation explains the higher-level phenomenon of a firm's network position.

Table 3 supports the hypothesis for all network measures except clustering. Already the first model, only taking into account the size of the firms, is capable of predicting degree centrality and number of triads. Adding the size of the patent portfolios improves the predictive power for degree, triads and especially closeness. All three measures become more correlated in model 3 when the distance benefit relationship in terms of overlap and its square are included. Clustering is not explained by any of the models. Probably due to the low number of triads in the network and the normalization by degree it is very difficult to predict. Nevertheless, we find that including the firm position in technological space, in model 2 the size dimension and in model 3 the structural dimension, helps to explain the firm position in network space.

	model 1	model 2	model 3
degree	0.51***	0.62***	0.64***
closeness	0.13	0.29***	0.32***
clustering	-0.07	-0.03	0
nb. triads	0.19**	0.3***	0.33***

Table 3: Pearson's correlation of observed and expected firm level network characteristics. Expected network characteristics are based on estimates of the random firm effects model by simple monte carlo estimation with 1000 draws. *,**,*** signify 5%, 1% and 0.1% rejection levels of a t-test of non-correlation.

4.1.3 Hypothesis 3

Hypothesis 3 proposes that network distributions depend on the firms' distribution in knowledge space. This hypothesis is supported but the effect is weak. The firms' technological characteristics, i.e. size of patent stock and overlap in IPC classes, improve explanation of the observed network distributions only slightly after the size of firms is taken into account.

Figure 1 compares observed with expected network distributions. The performance of the three models can be judged on how they improve the random model. The random model contains no firm information but only an intercept. This implies that all dyads have the same probability to be formed and the density of the observed network is met. [6] Interestingly, the process itself of dyad-wise partnering decision is likely to generate a centralized network with some clustering/triads. However, centralization and clustering/triads is lower than in the observed network and introduction of firm level information improves the expected distributions.

For all four network distributions we find a big improvement from the random model to model 1, where firm size is introduced. Introduction of the firm position in knowledge space, with model 2 and model 3, yields minor improvements relative to model 1.

The degree distribution is met best. The reason is that the regression estimates dyad formation and this is highly related to degree, which is simply the sum over all dyads formed by a firm. The other measures depend on more complex network structures. Closeness takes into account the whole network, triads the links between three firms and clustering the ratio of triads to degree. One important result is that these more complex network structures are predicted better by refinements of dyadic decision making other than including references to these structures. The distribution of firm size in the population together with the estimated heterophily of big and small firms implies higher closeness centralization and more triads/clustering than the random model does. However, improvements due to the distribution of firms in technological space is not that significant once firm size is controlled for.

4.2 Sensitivity Analysis

The results discussed above are based on the random effects logit, which assumes that firm specific effects are not correlated with other covariates. Besides the random effects logit, we estimated a fixed effects logit and compared the coefficients using a Hausman test (see Appendix B). The Hausman test shows that both models yield similar coefficient estimates, which justifies focusing on the random effects model.

A further problem might be inclusion of firms with few patents. Patents signal the technological position of firms. When a firm applies for few patents during the period of observation, the signal might not give the full range of technological fields a firm in fact covers. Then, the firm might be wrongly taken

[6] The random model is also known as Erdoes-Renyi model.

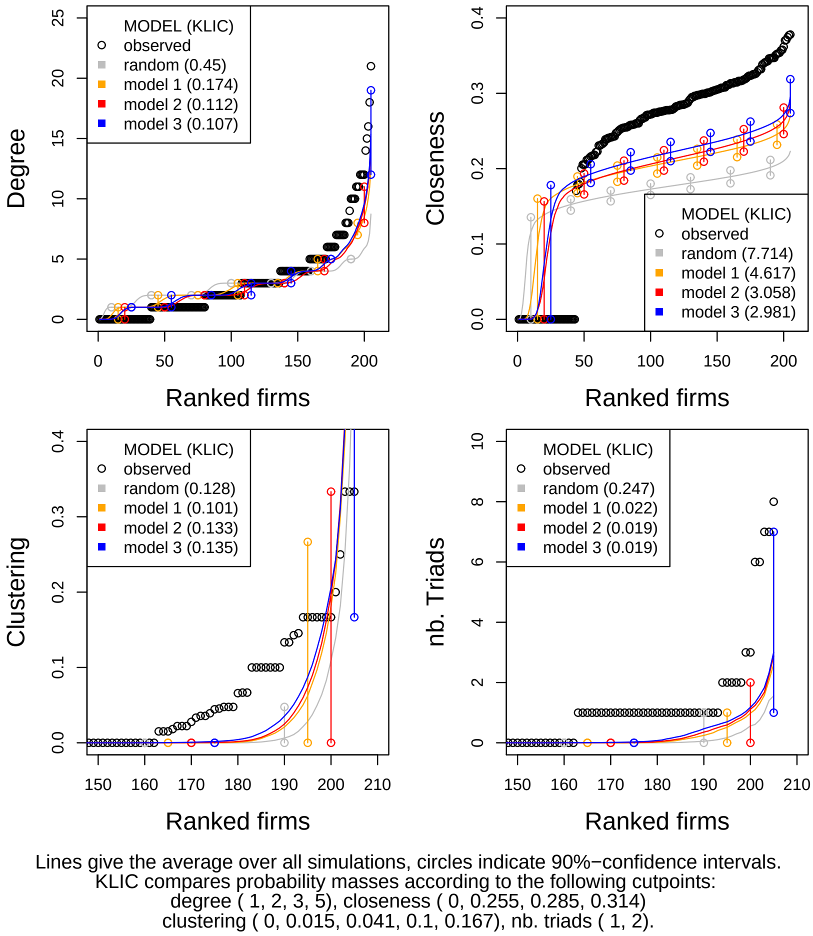

Figure 5: network level

as being at the boundary of technological space. The sensitivity of our results with respect to this problem is tested on firms having more than five patents and thus signal more reliably their position. This restriction does hardly change the coefficient estimates, whereas the level of significance of *overlap* and $overlap^2$ increases to 1% and 5% respectively. This is due to higher standard errors along with the reduced number of observations. Therefore, regression on the restricted sample supports hypothesis 1. Also, hypothesis 2 and hypothesis 3 are equal in magnitude and significance to the results already discussed above.

Finally, other distance measures than overlap have been applied. We repeated the analysis for the uncentered correlation of firms technology vector, introduced by (Jaffe, 1986, 1989), and the correlated revealed technological advantage (cRTA), introduced by (Soete, 1987; Patel and Pavitt, 1987). These measures have been developed for different reasons. As its predecessor, revealed comparative advantage, revealed technological advantage (RTA) has been applied to compare the relative specialization of countries. Jaffe (1986) aggregated all IPC classes into 49 technology fields to calculate the uncentered correlation for firms of various sectors.

Both distance measures are highly significant in regressing joint technological agreements. However, hypothesis one of an inverse-U-shaped relationship is not supported as the square of distance remains insignificant. Consequently, tests of hypothesis two and three are based on the preference for technological proximity as a local effect. Both hypothesis are supported similar in strength to those presented in the previous section. Thus, estimation with alternative distance measures strongly supports the existence of a local technological distance effect on the network. However, the exact shape of the local effect depends on the distance measure.

5 Discussion and Conclusion

5.1 Summary

This paper proposes a theoretic model of network formation and tests it empirically. In the theoretic model, firms are positioned in technological space. Two firms form a link whenever their technological distance is in some specified beneficial range. The model shows how the firms' distribution in technological space and the specification of the benefit-distance-range determines the alliancing decisions of all firm pairs. In the aggregate, the dyadic decisions imply a specific network structure and a specific network position for each firm. Variations on the nature of the technological space and the specification of the benefit-distance-range lead to qualitatively different network structures and network firm positions.

The empirical analysis confirms the relevance of technological distance for dyadic alliancing decisions and estimates this effect. Parallel to the theoretic model, the estimates yield expectations on the network structure and firm network positions. We find that the firms' position in the network is explained better when their position in technological space is taken into account. However, the network structure is largely defined by firm size. These results have been shown to be robust to more stringent estimation, sample restriction and alternative distance measures.

5.2 Implications on Theory

The model builds on the concept of optimal cognitive distance (Nooteboom et al., 2007), which is implied by the absorptive capacity of a firm (Cohen and Levinthal, 1990). Our empirical findings largely strengthen prior results that medium technological distance between firms is beneficial for research alliances (Mowery et al., 1998). The main contribution to this stream of research however is to show its implications on network formation, both theoretically and empirically.

The theoretic model continues the connection models of (Jackson and Wolinsky, 1996; Gilles and Johnson, 2000) by incorporating the concept of optimal cognitive distance. Contrary to previous work, we do not search for stability and efficiency configurations but focus on how certain benefit-distance specifications affect the network structure and firm positions.

This question is typical for social network analysis (Powell et al., 2005, e.g.). For answering it, we built on its strong empirical tradition (Hoff, 2003; Hunter et al., 2008) and, in order to account for dyadic independence, paralleled the typical fixed/random effects approach for panel data (Cameron and Trivedi, 2005).

Originally, social network analysis put socially motivated local effects on the forefront. This paper brings in a local effect which is motivated from research in knowledge economics (Cohen and Levinthal, 1990).

5.3 Implications on Research

The sensitivity analysis revealed that different distance measures result in different distance-benefit relationships. Whereas optimal cognitive distance has been attested by overlap of the firms IPC-vector, uncentered Correlation of IPC-vectors and correlated revealed technological advantage yielded a preference for proximity. Each distance measure has advantages over others in a certain context; i.e. is justified depending on the effects focused on and the sample chosen.

For example, we think that the overlap measure captures best the idea of absorptive capacity in research alliances, because it relates the technological fields new to the partner to the technological fields common to both firms. However, this is rather ad hoc and we are still missing sound micro-economic justifications for when to apply which distance measure. This also means to better connect patent based distance measures with other distance measures. This would improve measurement and help to gain a better intuition on patent based distance measures.

The measurement problem also becomes relevant, when extending our study to other industries. In the pharmaceutical industry patenting is current as appropriation, coordination and signaling device (Penin, 2005). We also found that bio-pharmaceutical patent classes differentiate among the firms, meaning that firms are active in different patent classes. Other industries might not be as convenient for measurement.

On the other hand, extending the analysis on other industries might prove fruitful. A strong feature of the pharmaceutical industry is its asymmetry between big and small firms. It seems that this prevented to see bigger effects of technological distance on the network structure. Research networks consisting of more equal firms might be structured more according to technological space. Furthermore, the theoretic model implies the potential existence of adverse benefit-distance effects. In other industries firms being central in knowledge space might have low network centrality. Such instances still need to be found.

Taken from a wider perspective, this paper tries to integrate different research streams to understand better network formation, as asked for by (Jackson et al., 2003). Why not make use of the rich literature on alliance formation in order to better explain the formation of networks? Recent advances in empirical methodology offer now a unified approach for such an endeavor.

5.4 Implications on Management

Big pharmaceutical firms have been and still are the central actors in the pharmaceutical industry. They are also central in the research network, whose structure is largely explained by firm size. This paper suggests that their centrality is not only due to their strong capabilities in financial and organizational aspects. In addition, technological diversification puts them in a central position in technological space and thus, opens up many opportunities for research alliances.

Emphasizing the effect of technological position on network position also yields management implications. In the short run, management can not freely envisage profitable network positions but is bounded by the firms technological endowment. This needs to be considered in the technology strategy of the firm.

Firms which focus on distant technological niches to reduce competitive pressure might find themselves isolated in the research network as well. Considering opportunities for cooperation besides unique technological qualification is crucial, because research alliances are important sources of financing and internal technological development.

5.5 Conclusion

This paper examines how the technological position of firms affects network formation. The theoretic model proposes that firms form research alliances depending on their mutual technological distance. In particular, firms do not consider the network structure or their position in the network. The theoretic model shows that such dyadic decision making is capable of producing different networks and putting firms in different network positions. The empirical study confirms that such a simple model indeed helps to explain industrial networks. Based on estimates of the dyadic alliance decision, we find that the network positions of firms are related to their position in knowledge space. The network structure seems to arise mainly due to size differences among the firms, which emphasizes the importance of financial and organizational interdependence in the pharmaceutical industry. In this way the paper sheds light on technological distance as a local effect of network formation.

References

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

Arora, A. and A. Gambardella, "Complementarity and External Linkages: The Strategies of the Large Firms in Biotechnology," *Journal of Industrial Economics* 38 (June 1990), 361–79.

Arundel, A. and I. Kabla, "What percentage of innovations are patented? empirical estimates for European firms," *Research Policy* 27 (1998), 127 – 141.

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

Cameron, A. and P. Trivedi, *Microeconometrics : methods and applications* (Cambridge Univ. Press, 2005).

Cantwell, J. and M. G. Colombo, "Technological and Output Complementarities: Inter-Firm Cooperation in Information Technology Ventures," *Journal of Management and Governance* 4 (2000), 117–147.

Cockburn, I., R. Henderson and S. Stern, "The Diffusion of Science-Driven Drug Discovery: Organizational Change in Pharmaceutical Research," (September 1999).

Cockburn, I. M. and R. M. Henderson, "Absorptive Capacity, Coauthoring Behavior, and the Organization of Research in Drug Discovery," *Journal of Industrial Economics* 46 (June 1998), 157–82.

Cohen, W. M. and D. A. Levinthal, "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly* 35 (1990), 128–152.

Cowan, R. and N. Jonard, "Network structure and the diffusion of knowledge," *Journal of Economic Dynamics and Control* 28 (June 2004), 1557–1575.

Eisenhardt, K. M. and C. B. Schoonhoven, "Resource-Based View of Strategic Alliance Formation: Strategic and Social Effects in Entrepreneurial Firms," *Organization Science* 7 (March-April 1996), 136–150.

EPO, "Worldwide Statistical Patent Database (PATSTAT)," (October 2008).

Fafchamps, M. and F. Gubert, "Risk Sharing and Network Formation," *American Economic Review* 97 (May 2007), 75–79.

Galambos, L. and J. L. Sturchio, "Pharmaceutical Firms and the Transition to Biotechnology: A Study in Strategic Innovation," *The Business History Review* 72 (1998), 250–278.

Gilles, R. P. and C. Johnson, "original papers : Spatial social networks," *Review of Economic Design* 5 (2000), 273–299.

Gilsing, V., B. Nooteboom, W. Vanhaverbeke, G. Duysters and A. van den Oord, "Network embeddedness and the exploration of novel technologies: Technological distance, betweenness centrality and density.," *Research Policy* 37 (2008), p1717 – 1731.

Henderson, R., L. Orsenigo and G. P. Pisano, *The Sources of Industrial Advantages*, chapter The Pharmaceutical Industry and the Revolution in Molecular Biology: Exploring the Interactions Between Scientific, Institutional and Organizational Change (Cambridge University Press, 1999), 267 – 311.

Hoff, P., "Random Effects Models for Network Data," *Dynamic Social Network Modeling and Analysis: Workshop Summary and Papers* (2003), 303–312.

Hunter, R. David, Goodreau, M. Steven, Handcock and S. Mark, "Goodness of Fit of Social Network Models," *Journal of the American Statistical Association* 103 (March 2008), 248–258.

Hunter, D. R., M. S. Handcock, C. T. Butts, S. M. Goodreau and M. Morris, "ergm: A Package to Fit, Simulate and Diagnose Exponential-Family Models for Networks," *J STAT SOFTW* 24 (12 2007), 1–29.

Jackson, M. O., G. Demange, S. Goyal and A. V. D. Nouwel, "A survey of models of network formation: stability and efficiency," in *In Group Formation in Economics: Networks, Clubs and Coalitions* (Cambridge University Press, 2003).

Jackson, M. O. and A. Wolinsky, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory* 71 (1996), 44 – 74.

Jaffe, A. B., "Technological Opportunity and Spillovers of R & D: Evidence from Firms' Patents, Profits, and Market Value," *The American Economic Review* 76 (1986), 984–1001.

———, "Characterizing the technological position of firms, with application to quantifying technological opportunity and research spillovers," *Research Policy* 18 (1989), 87 – 97.

McGee, D., "Making up Mind: The Early Sociology of Invention," *Technology and Culture* 36 (1995), 773–801.

Mowery, D. C., J. E. Oxley and B. S. Silverman, "Technological overlap and interfirm cooperation: implications for the resource-based view of the firm," *Research Policy* 27 (September 1998), 507–523.

Nooteboom, B., W. Van Haverbeke, G. Duysters, V. Gilsing and A. van den Oord, "Optimal cognitive distance and absorptive capacity," *Research Policy* 36 (September 2007), 1016–1034.

OECD, "The Measurement of Scientific and Technological Activities using Patent Data as Science and Technology Indicators: Patent Manual 1994," Technical Report, OECD, 1994.

———, "OECD Health Policy Studies: Pharmaceutical Pricing Policies in a Global Market," Technical Report, OECD, 2008a.

———, "OECD Patent Databases: Identifying Technology Areas for Patents," Technical Report, OECD, 2008b.

Orsenigo, L., F. Pammolli and M. Riccaboni, "Technological change and network dynamics: Lessons from the pharmaceutical industry," *Research Policy* 30 (March 2001), 485–508.

Patel, P. and K. Pavitt, "Is Western Europe losing the technological race?," *Research Policy* 16 (1987), 59 – 85.

Pavitt, K., "R&D, patenting and innovative activities : A statistical exploration," *Research Policy* 11 (1982), 33 – 51.

Penin, J., "Patents versus ex post rewards: A new look," *Research Policy* 34 (2005), 641 – 656.

Pfeffer, J. and P. Nowak, "Joint Ventures and Interorganizational Interdependence," *Administrative Science Quarterly* 21 (September 1976), 398 – 418.

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

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

Rosenkopf, L. and P. Almeida, "Overcoming Local Search Through Alliances and Mobility.," *Management Science* 49 (2003), p751 – 766.

Schmoch, U., F. Laville, P. Patel and R. Frietsch, "Linking Technology Areas to Industrial Sectors : Final Report to the European Commission, DG Research. In Corporation with Observatoire des Sciences et des Techniques, Paris and University of Sussex, Science and Policy Research Unit (SPRU), Brighton, United Kingdom," (2003).

Schoenmakers, W. and G. Duysters, "Learning in strategic technology alliances," *Technology Analysis & Strategic Management* 18 (2006), 245–264.

Shan, W., G. Walker and B. Kogut, "Interfirm Cooperation and Startup Innovation in the Biotechnology Industry," *Strategic Management Journal* 15 (1994), 387–394.

Singh, J., "Collaborative Networks as Determinants of Knowledge Diffusion Patterns," *Management Science* 51 (2005), 756–770.

Soete, L., "The impact of technological innovation on international trade patterns: The evidence reconsidered," *Research Policy* 16 (1987), 101 – 130.

Stuart, T. E., "Network positions and propensities to collaborate," *Administrative Science Quarterly* 43 (1998), 668–698.

Train, K. E., *Discrete Choice Methods with Simulation* (Cambridge University Press, 2003).

van Duijn, M. and J. Vermunt, "What Is Special About Social Network Analysis?," *Methodology* 2 (2006), 2–6.

Verspagen, B., T. van Moergastel and M. Slabbers, "MERIT concordance table: IPC - ISIC (rev.2)," Technical Report, MERIT Research Memoranda, 1994.

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

Watts, D. J. and S. H. Strogatz, "Collective dynamics of 'small-world' networks.," *Nature* 393 (June 1998), 440–442.

A Derivation of Network Measures

A.1 degree centrality

The degree of a firm can be derived by integrating the neighbors over the neighborhood range of the firm. In the unbounded knowledge space all firms are in the same situation. The agent in the origin 0 forms links with all partners $j \in [a, a+b] \cup [-a, -a-b]$. Assuming a neighborhood density of one, the degree for any firm i becomes $2 \int_a^{a+b} 1\, dj = 2b$ and the degree distribution of the graph is a point mass at $2b$.

In the bounded knowledge space the degree of firm j depends on its position and the relative sizes of a and $a+b$. We can calculate the degree of the right-hand links for three kinds of agents j:

$$\begin{array}{rl} \text{for} j \in [0, 1-a-b] & degree_j = \int_{j+a}^{j+a+b} 1\, di = b \\ \text{for} j \in [1-a-b, 1-a] & degree_j = \int_{j+a}^{1} 1\, di = 1-j-a \\ \text{for} j \in [1-a, 1] & degree_j = \int_{1}^{1} 1\, di = 0 \end{array}$$

The first kind of agent realizes all links in the right-hand benefit range, the second is partly restricted by the boundary 1 and the third does not realize any right hand neighbors. Similarly, the left-hand links for any agent j are determined. The degree is then the sum over right- and left-hand links for agent j. Given the span of the benefit distance range this results in the two cases given in the main text.

A.2 closeness centrality

Closeness centrality is defined as one divided by the average shortest path between a vertex i and all other vertices reachable from it:

$$closeness_i = 1/\left(\frac{1}{N}\sum_{j=1, j\neq i}^{N} d_{ij}\right)$$

where d_{ij} is the shortest path (i.e. the minimum number of links) connecting two vertices i and j in the network.

When knowledge space is unbounded the network of firms is infinitely large. Therefore, each firm has infinitely long paths which implies a closeness coefficient of 0 for all nodes.

To prove this formally, define the average path length of agent 0 $path_0$ including agents reached in s_r steps to the right:

$$path_0 = \frac{1}{s_r(a+b)}\left[\sum_{j=1}^{s_r} j(a+b)+a\right] = \sum_{j=1}^{s_r}\frac{j}{s_r} + \frac{a}{s_r(a+b)}$$

if $s_r \to \infty$, the second term $\frac{a}{s_r\,(a+b)} \to 0$ and we need to consider only the first term $\sum_{j=1}^{s_r}\frac{j}{s_r}$:

$$\begin{aligned}
\sum_{j=1}^{s_r}\tfrac{j}{s_r} &= \sum_{j=1}^{s_r}\tfrac{s_r-j}{s_r}\\
&= \sum_{j=1}^{s_r}\left(1-\tfrac{j}{s_r}\right)\\
&= \sum_{j=1}^{s_r}1-\sum_{j=1}^{s_r}\tfrac{j}{s_r}\\
&= \sum_{j=1}^{s_r}1/2\\
&= s_r\;1/2
\end{aligned}$$

The first equation reorders the sum. The last equation shows that the average path length to right agents goes to infinity with s_r. Because all agents are in the same situation, closeness is 0 for all of them.

A general formula for closeness may be derived for the bounded knowledge space. Closeness of a firm i is derived by averaging the shortest paths to all firms on its left and on its right. Besides the first and the last step, the mass of firms covered in one step again is $(a+b)$. Because the space is bounded, a firm at position i has $\lfloor i/(a+b)\rfloor = s_l$ full steps to the left and $\lfloor (1-i)/(a+b)\rfloor = s_r$ steps to the right. Taking into account the first and the last step, the sum of shortest paths to agents on the left becomes:

$$paths_i^l = \begin{cases} \sum_{j=1}^{s_l} j(a+b)+(s_l+1)(i-s_l(a+b))+a & \text{if } i>2a \mid a<i<1-a\\ 2i & \text{if } i<a \;\&\; i<1-a\end{cases}$$

where the first term in the first line sums up the mass of agents covered in each full step, weighted by the path length j. The second term adds the last step which covers $(i-s_l(a+b))$ agents and the last term corrects for the first step. The second line takes into account the case, where no alliances with firms on the left are beneficial for i and thus left firms can only be reached via partners from the right. As long as $1-a>i$ (there are partners to the right) and $b>a$ all left firms can be reached in the second step via the closest firm to the right $(i{+}a)$.

The sum of all paths to agents on the right is similar:

$$paths_i^r = \begin{cases} \sum_{j=1}^{s_r} j(a+b) + (s_r+1)(1-(i+s_r(a+b))) + a & \text{if } i < 1-a \\ 2(1-i) & \text{if } i > 1-a \text{ and } i > a \end{cases}$$

Both together, $paths_i^l$ and $paths_i^r$, yields a function of the average path of firm i depending on its position i in knowledge space. Applying these more general formulas to the two cases where firms in the center of the knowledge space are more (case 1) or less (case 2) restricted by the boundaries yields the result described in the main text.

A.3 clustering coefficient

Define the neighborhood of node i as the set of all neighbors $N_i = \{j|d_{ij} = 1\}$ where d_{ij} is the distance between i and j in network space. In an undirected graph of size $|N_i|$ there could exist $\frac{|N_i|(|N_i|-1)}{2}$ links. If we write the existing links among neighbors as $\{e_{jk}\}$ where $j, k \in N_i$, the clustering coefficient for node i becomes $clustering_i = \frac{2|\{e_{jk}\}|}{|N_i|(|N_i|-1)}$.

In the unbounded knowledge space one might again consider agent 0 as representative agent. Consider $N_0^R = \{j > 0 \in N_0\} = [a, a+b]$, that is agent 0's right neighbors. For any agent $j \in N_0^R$, the mass of links to $k \in N_0^R$ is $N_j^R \cap N_0^R$. To avoid the double counting due to bi-directional links, we can consider only links of j to the right of j located in N_0^R. This will be the set of agents lying between the leftmost neighbor of j and the right-most neighbor of 0. That is, $[j+a, a+b]$ for $j \leq b$. For any agent j, $N_j^R \cap N_0^R = \int_{j+a}^{a+b} 1dk = b - j$. Integration over all nodes $j \in N_0^R$ possibly contributing to clustering yields $N_{N_0^R}^R \cap N_0^R = \int_a^b (b-j)dj = 1/2(b-a)^2$. Since distance is symmetric around the originating agent, the neighborhood to the left of agent 0 is identical in this respect: $N_{N_0^L}^L \cap N_0^L = N_{N_0^R}^R \cap N_0^R$. It remains to find the neighbors to the left who are linked to neighbors on the right: $N_{N_0^L}^R \cap N_0^R$. Again, for agent $j \in N_0^L$, this is from the left-most neighbor of 0 to the right-most neighbor of j, i.e. the range $[a, j+a+b]$ yielding $N_0^R \cap N_j^R = \int_a^{j+a+b} 1dk = j + b$. Integration over the left neighborhood gives: $N_0^R \cap N_{N_0^L}^R = \int_{-b}^{-a} (j+b)dj = 1/2(b-a)^2$. Thus, the total number of links among neighbors of node 0 is $3/2(b-a)^2$ and we can state directly:

Proposition 4 *For each agent, the clustering coefficient is* $\frac{3}{4}\frac{(b-a)^2}{b^2}$.

In the bounded knowledge space, the general principle of clustering is identical to the infinite case. However the intervals of integration change depending on the position of the agents. It is convenient first to derive the clustering for general boundaries and then inserting the boundaries for different cases.

Consider agent i. Define his left and his right neighborhood as $N_i^L = [L_{i,M}, L_{i,m}]$ and as $N_i^R = [R_{i,m}, R_{i,M}]$ respectively. If an agent $j \in N^{R_i}$ has neighbors in N_i^R, this contributes to i's clustering coefficient. As before, if we consider only j's right neighbors, we avoid double-counting links. The overlap $N_i^R \cap N_j^R$ runs from the left-most right neighbor of j to the rightmost right neighbor of i: $[R_{j,m}, R_{i,M}]$. This we integrate over $j \in N_i^R$:

$$E_i^R = \int_{R_{i,m}}^{R_{i,M}} max(0, R_{i,M} - R_{j,m}) dj$$

Similarly regarding i's left neighborhood:

$$E_i^L = \int_{L_{i,M}}^{L_{i,m}} max(0, L_{j,m} - L_{i,M}) dj$$

Finally, there could be left-hand neighbors of i who are connected to right-hand neighbors of i:

$$E_i^{LR} = \int_{L_{i,M}}^{L_{i,m}} max(0, R_{j,M} - R_{i,m}) dj$$

The sum is the number of links among neighbors of agent i, $E_i^N = E_i^R + E_i^L + E_i^{LR}$. Whereas in the unbounded case integration was simply over the neighborhood, in the bounded knowledge space the effect of the boundaries need to be taken into account. The boundaries might have i) no effect on the neighborhood (NO), ii) restrict the neighborhood (R) or iii) completely prevent a neighborhood (P). In order to calculate the clustering coefficient of an agent i, the effect of the 0 boundary on the left-hand neighborhood as well as the effect of the 1 boundary on the righ-hand neighborhood needs to be considered. Because of symmetry it suffices to distinguish the combinations (NO, NO), (R,NO), (P,NO), (R,R), (P,R).

(NO, NO) If no neighborhood is affected, the result of the unbounded knowledge space applies and the clustering coefficient of agent i becomes $\frac{3}{4}\frac{(b-a)^2}{b^2}$.

(R, NO) The left neighborhood is restricted if $i \in [a, a+b]$. Additionally, i needs to be $\in [0, 1-a-b]$ to be not restricted in the right neighborhood. For this to be the case, both sets need to overlap, which implies that $i \in [a, a+b]$ if $1 > 2(a+b)$ or $i \in [a, 1-a-b]$ if $1 < 2(a+b)$. In both cases the clustering coefficient is the same, because the boundaries for integration are the same. We derive the clustering coefficient for $i \in [a, a+b]$ in the following way: since i is not restricted on the right-hand, the right-neighborhood of i is $N_i^R = [i+a, i+a+b]$. Restriction on the left-hand gives $N_i^L = [0, i-a]$. An agent of the right neighborhood $j_r \in N_i^R$ might be restricted or not:

$$N_{j_r}^R = \begin{cases} [j+a, j+a+b] & \text{if } j < 1-a-b \\ [j+a, 1] & \text{if } j > 1-a-b \end{cases}$$

Agents of the left neighborhood are necessarily restricted or prevented on the left side

$$N_{j_l}^L = \begin{cases} [0, j-a] & \text{if } j > a \\ \oslash & \text{else} \end{cases}$$

, whereas no restriction will be on the right side $N_{j_l}^R = [j+a, j+a+b]$. To derive the number of links in the neighborhood, it suffices to fill in the boundaries of the specific case. For (R,NO) we get:

$$\begin{aligned} E_i^R &= \textstyle\int_{R_{i,m}}^{R_{i,M}} max(0, R_{i,M} - R_{j,m})\, dj \\ E_i^R &= \textstyle\int_{i+a}^{i+a+b} max(0, i+a+b-(j+a))\, dj \\ E_i^R &= \textstyle\int_{i+a}^{i+b} (i+b-j)\, dj \end{aligned}$$

where the last equation shifts the boundary of integration to simplify the inner expression and $b > a$ is assumed. Similarly, the links among left-hand neighbors:

$$\begin{aligned} E_i^L &= \textstyle\int_{L_{i,M}}^{L_{i,m}} max(0, L_{j,m} - L_{i,M})\, dj \\ E_i^L &= \textstyle\int_0^{i-a} max(0, j-a-0)\, dj \\ E_i^L &= \textstyle\int_a^{i-a} (j-a)\, dj \end{aligned}$$

for agents with $i - a > a$ or $i > 2a$, otherwise $E_i^L = 0$. Links among left and right-hand neighbors are:

$$\begin{aligned} E_i^{LR} &= \textstyle\int_{L_{i,M}}^{L_{i,m}} max(0, R_{j_l,M} - R_{i,m})\, dj \\ E_i^{LR} &= \textstyle\int_0^{i-a} max(0, j+a+b-(i+a))\, dj \\ E_i^{LR} &= \textstyle\int_0^{i-a} (j+b-i)\, dj \qquad \text{if } i < b \\ E_i^{LR} &= \textstyle\int_{i-b}^{i-a} (j+b-i)\, dj \qquad \text{if } i > b \end{aligned}$$

Calculation of the integrals yields: $E_i^R = \frac{1}{2}(b-a)^2$, $E_i^L = \frac{1}{2}(i-2a)^2$ for $i > 2a$ and $E_i^L = 0$ for $a < i < 2a$ and $E_i^{LR} = \frac{1}{2}(i-a)(b-a)$ for $i < b$ and $E_i^{LR} = \frac{1}{2}(b-a)^2$ for $i > b$. The total number of links among neighbors of an agent i, which is partly restricted on one side is therefore always smaller than for the unrestricted agent. However, the clustering coefficient norms the number of realized links by the number of potential links, which is $1/2(i-a+b)^2$. Both, the number of links and the potential number of links, are functions of the position in the knowledge space ($f(i)$). Because the number of potential links increases faster with i moving away from the boundary, the clustering coefficient decreases along the way. (To see this compare the slopes of both functions for specific i's.) Thus, the higher the restriction on one side, the higher the clustering coefficient.

(P,NO) When agent i has no left-hand neighborhood but a right-hand neighborhood over the whole benefit range, we can state directly $E_i = E_i^R = \frac{1}{2}(b-a)^2$, which yields a clustering coefficient of $\frac{(b-a)^2}{b^2}$. This case is at the extreme of the previous case. Thus, as the restriction of the neighborhood due to a boundary increases, the clustering coefficient increases from $\frac{3}{4}\frac{(b-a)^2}{b^2}$ up to $\frac{(b-a)^2}{b^2}$. The number of links among the neighborhood of course decreases from $3/2(b-a)^2$ to $(b-a)^2$.

(P,RE) agents with no left-hand neighborhood and a restricted right-hand neighborhood might only occur in case 1 of figure 2, where $1 < 2a + b$. Again, it suffices to consider the right-hand contribution of links:

$$\begin{array}{rcl} E_i^R & = & \int_{R_{i,m}}^{R_{i,M}} max(0, R_{i,M} - R_{j,m})\, dj \\ E_i^R & = & \int_{i+a}^{1} max(0, 1-(j+a))\, dj \\ E_i^R & = & \int_{i+a}^{1-a} (1-a-j)\, dj \\ E_i^R & = & \frac{1}{2}(1-2a-i)^2 \end{array}$$

Because the size of the neighborhood is $1-a-i$, the clustering coefficient becomes $(1-2a-i)^2/(1-a-i)^2$, which is a decreasing function with i given $a > 0$. In case 1 of figure 2, clustering decreases until $i = a$ to $(1-3a)^2/(1-2a)^2$. This might be more or less than the unbounded clustering coefficient $3/4(b-a)^2/b^2$, depending on the specification of a and b. When $i > a$, the left hand side needs to be considered and we move to the next regime (R,R).

(R,R) For a restriction on the left-hand to occur i needs to be $\in [a, a+b]$, for the right-hand $i \in [1-a-b, 1-a]$. This will only happen if the diameter of the knowledge space is smaller than two times the benefit range, i.e. if $1 < 2(a+b)$. A restriction on both sides might occur in both cases displayed in figure 2. In case 1, where $a > 1-a-b$, $i \in [a, 1-a]$. Case 2, with $a < 1-a-b$ implies that $i \in [1-a-b, a+b]$. However, the boundaries implied for the neighborhood of the focal agent i are the same in both cases, with $N_i^R = [i+a, 1]$ and $N_i^L = [0, i-a]$. For agent j_r and j_l out of the right and left neighborhood respectively, the neighborhood range is:

$$N_{j_r}^R = \begin{cases} [j+a, 1] & \text{if } j > 1-a \\ \oslash & \text{else} \end{cases}$$

$$N_{j_l}^R = \begin{cases} [j+a, j+a+b] & \text{if } j < 1-a-b \\ [j+a, 1] & \text{if } 1-a-b < j < 1-a \end{cases}$$

$$N_{j_l}^L = \begin{cases} [0, j-a] & \text{if } j > a \\ \oslash & \text{else} \end{cases}$$

As before the contribution of the right- and left-hand neighborhood alone and the links from the left to the right-hand neighborhood are:

$$
\begin{aligned}
E_i^R &= \int_{R_{i,m}}^{R_{i,M}} max(0, R_{i,M} - R_{j,m})\, dj \\
E_i^R &= \int_{i+a}^{1} max(0, 1 - (j + a))\, dj \\
E_i^R &= \int_{i+a}^{1-a} (1 - a - j)\, dj \\
E_i^R &= \tfrac{1}{2}(1 - 2a - i)^2
\end{aligned}
$$

for agents with $1 - a > i + a$ or $i < 1 - 2a$, otherwise $E_i^R = 0$. Similarly, the links among left-hand neighbors:

$$
\begin{aligned}
E_i^L &= \int_{L_{i,M}}^{L_{i,m}} max(0, L_{j,m} - L_{i,M})\, dj \\
E_i^L &= \int_{0}^{i-a} max(0, j - a - 0)\, dj \\
E_i^L &= \int_{a}^{i-a} (j - a)\, dj \\
E_i^L &= \tfrac{1}{2}(i - 2a)^2
\end{aligned}
$$

for agents with $i - a > a$ or $i > 2a$, otherwise $E_i^L = 0$. Links among left and right-hand neighbors are:

$$
E_i^{LR} = \int_{L_{i,M}}^{L_{i,m}} max(0, R_{j_l,M} - R_{i,m})\, dj
$$

To insert the boundaries, one needs to distinguish the cases, where the contributing part of the left-hand neighborhood is restricted ($i < b$) and the contributing part of the right-hand neighborhood is restricted ($i > 1 - b$). Each combination in principle is possible:

$$
E_i^{LR} = \begin{cases}
\int_0^{i-a} (j + b - i)\, dj & \text{if } i < 1 - b \ \&\ i < b \\
\int_{i-b}^{i-a} (j + b - i)\, dj & \text{if } i < 1 - b \ \&\ i > b \\
\int_0^{1-a-b} (j + b - i)\, dj + \int_{1-a-b}^{i-a} (1 - a - i)\, dj & \text{if } i > 1 - b \ \&\ i < b \\
\int_{i-b}^{1-a-b} (j + b - i)\, dj + \int_{1-a-b}^{i-a} (1 - a - i)\, dj & \text{if } i > 1 - b \ \&\ i > b
\end{cases}
$$

$$
E_i^{LR} = \begin{cases}
(b - i)(i - a) + 1/2(i - a)^2 & \text{if } i < 1 - b \ \&\ i < b \\
1/2(b - a)^2 & \text{if } i < 1 - b \ \&\ i > b \\
-1/2(1 - a - b)^2 + (1 - a - i)(i - a) & \text{if } i > 1 - b \ \&\ i < b \\
(b - (1 - i))((1 - i) - a) + 1/2((1 - i) - a)^2 & \text{if } i > 1 - b \ \&\ i > b
\end{cases}
$$

To see, how clustering changes with i moving along the line, we might look at the derivations. The links contributed solely by the left and right hand side change with $\partial(E_i^R + E_i^L)/\partial i = 2i - 1$. Thus, the link contribution is first decreasing with i until $i = 1/2$ and then increasing again. For the left to right changes in link contribution, we find:

$$
\frac{\partial E_i^{LR}}{\partial i} = \begin{cases}
b - i & \text{if } i < 1 - b \ \&\ i < b \\
0 & \text{if } i < 1 - b \ \&\ i > b \\
1 - 2i & \text{if } i > 1 - b \ \&\ i < b \\
1 - b - i & \text{if } i > 1 - b \ \&\ i > b
\end{cases}
$$

Consider case 1 in figure 1. In the regime, where the agent is restricted on both neighborhoods, we move from a to $1-b$ where $i < 1-b$ and $i < b$. Thus, the contribution changes with a slope ($\partial E_i/\partial i = (2i-1)+(b-i) = i+b-1$. Since $i < 1-b \Rightarrow i+b-1 < 0$, the contribution is decreasing. In the range of $1-b$ to b, where $i > 1-b$ and $i < b$ changes level off and the links among neighbors remain constant. Beyond b the links increase again as they decreased before because of symmetry.

In case 2 in figure 2 the same dynamic is happening. Since we come from the (NO,R) regime where the number of links decreases and in the new regime the links decrease even more with the size of the neighborhood being stable, the clustering coefficient is smaller in the whole regime with some valley in the middle.

Hence, the main qualitative result is that the boundaries reduce the neighborhood of a firm if its left and/or right benefit-range is at least partly outside the boundaries. Independent of the setting of the benefit-range, i.e. the values of a and b, a firm closer to one boundary will always have a higher clustering coefficient than a firm closer to the center. This is because the number of realized links among neighbors increases always slower than the number of potential links when firms move away from the boundary.

B Random and Fixed Effects

B.0.1 models and estimation

For the logit model with firm specific effects the conditional probability of alliance formation is:

$$p_{ij} \;=\; Pr[y_{ij}=1|x_{ij},a_i,a_j] = \frac{\exp(x'_{ij}\beta +\, a_i \,+\, a_j)}{1+\exp(x'_{ij}\beta +\, a_i \,+\, a_j)}\;,$$

where p_{ij} is the probability that firm i and j form an alliance (i.e. $y_{ij}=1$), x_{ij} is a vector of dyadic-covariates, and a_i and a_j are firm specific effects. The probability mass function is written as:

$$f(y_{ij}|x_{ij},a_i,a_j) = p_{ij}^{y_{ij}}(1-p_{ij})^{1-y_{ij}}$$

The model assumes that firm specific effects are the only source of dependence and hence, given a_i and a_j, the dyadic observations are assumed to be independent.

Estimation of the fixed effects model is easily done with introduction of firm dummies. The assumption is that whereas the number of firms is fixed ($n \to$ constant) the number of observations goes to infinity ($n(n-1)/2 \to \infty$) which is not true but approximately given. Then, estimation is feasible via

Maximum Likelihood.

However, direct estimation of firm dummies is inefficient. To estimate the more efficient random effects model, the firm specific effects are integrated out. We do this with a direct monte carlo simulator under the assumption that the a_i are i.i.d. from a normal distribution $N(0, \sigma^2)$. The average of S draws yields the simulated probability, now conditional on known (simulated) firm specific effects and the variance of the distribution to be optimized:

$$\hat{f}(y_{ij}|x_{ij}, a_{is}, a_{js}, \sigma) = \frac{1}{S} \sum \frac{\exp(x'_{ij}\beta + \sigma a_{is} + \sigma a_{js})}{1 + \exp(x'_{ij}\beta + \sigma a_{is} + \sigma a_{js})} ,$$

where the a_{is} are i.i.d. draws from $N(0, 1)$ and transformed to firm specific effects by multiplication with the parameter σ. The simulated densities enter the maximum simulated likelihood estimator, which maximises:

$$lnL(\beta) = \sum \ln \hat{f}[y_{ij}|x_{ij}, a_{is}, a_{js}, \sigma]$$

over all firm-pairs. As long as $S, N \to \infty$ and $\sqrt{N}/S \to 0$, the single simulations (one draw) are unbiased and the usual assumptions for likelihood estimation apply, the estimator has a limit normal distribution with

$$\sqrt{N}(\hat{\theta}_{MSL} - \theta_0) \xrightarrow{d} N[0, A^{-1}(\theta_0)]$$

with

$$A(\theta_0) = -plim \left[N^{-1} \sum \frac{\delta^2 \ln f(y_{ij}|x_{ij}, \theta)}{\delta\theta\delta\theta} \right]$$

see (CameronTrivedi2005, p.393ff). The variance matrix is needed to derive confidence intervals and can be estimated in various ways. We choose the simplest estimator which is the BHHH estimate for the information matrix (CameronTrivedi2005, p.393ff).

The simulated likelihood is estimated with the iterative BroydenFletcher-GoldfarbShanno (BFGS) method. Here, as in other optimization procedures (e.g. Newton-Raphson, BHHH) the direction of the steps towards the optimum is given by the gradient in the current step and the size of the step is determined by the slope of the likelihood-function. The difference is that whereas other approaches use information for the slope only given by the current position (for Newton-Raphson the Hessian matrix, for BHHH the information matrix), BFGS determines the slope of the likelihood function by differences of the gradient caused by non-marginal position changes. This gives speed advantages in non-simple environments (Train, 2003, p.201) as can be expected for our problem.

For optimization we use the optim function in the R-stats-package to which we provide the simulated likelihood function:

$$lnL(\beta) = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N(N-1)/2} \ln \frac{1}{S} \sum_{s=1}^{S} \left(\frac{\exp(.)}{1+\exp(.)} \right)^{y_{ij}} \left(\frac{1}{1+\exp(.)} \right)^{1-y_{ij}},$$

where $\exp(.) = x'_{ij}\beta + \sigma a_i + \sigma a_j$. To ensure a positive variance σ, we optimize $log(\sigma)$ which results in a log-normal distribution for its standard error. In order to increase estimation speed, we derive the gradient of the MSL estimator. Because there is no principal difference between β and σ in the following, we combine them to θ with indicators for firm specific effects also incorporated in x_{ij}.

$$\frac{\delta lnL(\theta)}{\delta\theta} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N(N-1)/2} \left(\frac{\delta \frac{1}{S} \sum_{s=1}^{S} \left(\frac{\exp(.)}{1+\exp(.)} \right)^{y_{ij}} \left(\frac{1}{1+\exp(.)} \right)^{1-y_{ij}} / \delta\beta}{\frac{1}{S} \sum_{s=1}^{S} \left(\frac{\exp(.)}{1+\exp(.)} \right)^{y_{ij}} \left(\frac{1}{1+\exp(.)} \right)^{1-y_{ij}}} \right)$$

because $\frac{\delta \ln f_{ij}}{\delta\theta} = \frac{\delta f_{ij}/\delta\theta}{f_{ij}}$ and after some calculation

$$\frac{\delta lnL(\theta)}{\delta\theta} = \frac{2}{N(N-1)} \sum_{i=1}^{N} \sum_{j=i+1}^{N(N-1)/2} \left(\frac{\frac{1}{S} \sum_{s=1}^{S} \left(\left(y_{ij} x_{ij} \frac{\exp(.)}{(1+\exp(.))^2} \right) - \left((1-y_{ij}) x_{ij} \frac{\exp(.)}{(1+\exp(.))^2} \right) \right)}{\frac{1}{S} \sum_{s=1}^{S} \left(\frac{\exp(.)}{1+\exp(.)} \right)^{y_{ij}} \left(\frac{1}{1+\exp(.)} \right)^{1-y_{ij}}} \right)$$

Comparison of fixed and random effects models is based on the simplified version of the Hausman test. Under the assumption that the random effects estimate is fully efficient the covariances among the coefficients of the two models equal the variance of the efficient model coefficients (Cameron and Trivedi, 2005, p.272). This allows for separate estimation of both models, which simplifies the Hausman test.

B.1 results

In the fixed effects model the firm dummy controls for the overall alliance activity of the firm. If for a firm no alliance is observed, the dummy coefficient takes on minus infinity and hence is not defined. Therefore, a comparison of fixed and random effects can only be done on the restricted set of 166 firms, which have alliance partners in the network.

Table 4 gives the results of the random and fixed effects model as well as the Hausman test, which compares their coefficients. Except for *absDiffLnPC*, for no coefficient the null hypothesis of random and fixed effects estimation being equal can be rejected. This justifies to base analysis in the main text on the random effects estimates.

The random effects model coefficients *overlap* and *overlap*^{2} are still significant when estimated on the restricted firm sample. However, compared to the estimation on the complete sample magnitude decreases (see table 4.1.1 in the main text). Figure 3 reveals the reason: many firms with no alliance partners are at the boundary of the knowledge space; which supports hypothesis one.

In the fixed effects model *overlap* and *overlap*^{2} are not significant. Although the Hausman test confirms that coefficients are similar to the random effects estimation, increasing standard errors prevent significance. This effect can be largely attributed to the efficiency loss due to firm dummy estimation. Therefore the fixed effects estimation does not necessarily refuse Hypothesis one.

The heterophily of big and small firms in terms of patent counts and employees is confirmed in both models. Although, the coefficient capturing the difference in the number of patents changes significantly, it remains positive and significant even in the fixed effects model.

	random effects	fixed effects	H-value	$Pr(>\|H\|)$
intercept	-7.39*** (0.439)	–	–	–
overlap	3.44*** (1.01)	2.38 (1.83)	0.48	0.49
overlap^2	-1.81* (1.061)	-1.79 (1.74)	0.00	0.99
absDiffLnPC	0.25*** (0.043)	0.14** (0.061)	5.78	0.02
sumLnPC	0.05 (0.033)	0 (0.911)	0.00	0.96
absDiffLnEmpl	0.22*** (0.032)	0.2*** (0.038)	1.50	0.22
sumLnEmpl	0.04 (0.025)	-0.49 (1.016)	0.27	0.60
firmDummies	no	yes	–	–
σ^2	0.53 (0.111)	–	–	–
AIC	2907.01	2956.75	–	–

Table 4: Random and fixed effects models compared using the Hausman test. 13695 firm pair observations from crossing 166 firms. Standard errors in brackets; *,**,*** signify 5%, 1% and 0.1% rejection levels of significance. Hausman test null hypothesis: coefficients of random and fixed effects estimations are equal. $Pr(> |H|)$ is significance level of rejection of equality of coefficients derived from the chi-square distributed H-value with one degree of freedom.

In total, the comparison of random and fixed models justifies the focus on the random effects model and further supports hypothesis one.