// NO.21-049 | 06/2021

DISCUSSION PAPER

// MILAD ABBASIHAROFTEH, JAN KINNE, AND MIRIAM KRÜGER

The Strength of Weak and Strong Ties in Bridging Geographic and Cognitive Distances

The strength of weak and strong ties in bridging geographic and cognitive distances

Milad Abbasiharofteh[1,2*], Jan Kinne[3,4], Miriam Krüger[5]
[1] Department of Human Geography and Spatial Planning, Utrecht University, the Netherlands
[2] Department of Structural Change, Leibniz Institute of Agricultural Development in Transition Economies; 06120 Halle, Germany
[3] Department of Economics of Innovation and Industrial Dynamics, Leibniz Centre for European Economic Research (ZEW); 68161 Mannheim, Germany
[4] istari.ai UG (haftungsbeschränkt); 68199, Mannheim, Germany
[5] Chair of Innovation Economics, Faculty of Economics and Management, Berlin Institute of Technology; 10587 Berlin, Germany

* Corresponding author: m.abbasiharofteh@uu.nl

Abstract
The proximity framework has attracted considerable attention in a scholarly discourse on the driving forces of knowledge exchange tie formation. It has been discussed that too much proximity is negatively associated with the effectiveness of a knowledge exchange relation. However, little is known about the key factors that trigger the formation of the boundary-spanning knowledge ties. Going beyond the "dyadic" perspective on proximity dimensions, this paper argues that the key factor in bridging distances may reside at the "triadic" level. We build on the notion of "the strength of weak ties" and its recent development by investigating the innovative performance and relations of more than 600,000 German firms. We explored and extracted information from the textual and relational content of firms' websites by using machine learning techniques and hyperlink analysis. We thereby proxied the innovative performance of firms using a deep learning text analysis approach and showed that the triadic property of bridging dyadic relations is a reliable predictor of firms' innovativeness. Relations embedded in cliques (i.e., strong ties) that connect cognitively distant firms are more strongly associated with firms' innovation, whereas inter-regional relations connecting different parts of a network (i.e., weak ties) are positively associated with firms' innovative performance. Also, the results suggest that a combination of strong inter-community and weak inter-regional relations are more positively related with firms' innovativeness compared to the combination of other relation types.

Keywords: weak and strong ties, proximity, knowledge exchange, innovation, web mining, natural language processing

JEL codes: C81, D83, L14, O31

1. Introduction

The ever-increasing complexity of knowledge has stimulated firms to tap into external knowledge pools by creating collaborative relationships to seek required knowledge to innovate, and overcome technical and organizational bottlenecks (Broekel 2019; van der Wouden 2020). Investigating the driving forces behind the creation of knowledge exchange ties (also known as knowledge sourcing) has been of particular interest in economic geography (Clark et al. 2018). Economic geographers took the pioneering work of Marshall (1920) as a point of departure to show that colocation of firms increases the likelihood of knowledge transfer (Duranton and Puga 2004). More recently, the proximity framework, backed by a plethora of empirical studies, has become the staple of a scholarly discourse on the driving forces of knowledge tie formation in economic geography (Boschma 2005; Torre and Rallet 2005; Boschma and Frenken 2010; Balland, Boschma, and Frenken 2015; Balland, Boschma, and Frenken 2020). The proximity perspective proposes that geographical proximity should be regarded alongside other proximity dimensions (e.g., cognitive proximity) and that overlooking such dimensions may lead to overstating the effect of geographical proximity in triggering the creation of knowledge ties (Singh 2005; Breschi and Lissoni 2009; Balland, Boschma, and Frenken 2020).

Being close in one or several proximity dimensions increases the likelihood of the formation of knowledge transfer relations. Yet, interactions with, among others, geographically and cognitively close collaborators are associated with minimum benefits for learning and innovative capability in the long run due to an increased degree of redundancy (Boschma and Frenken 2010; Balland, Boschma, and Frenken 2020). Thus, scholars argue that firms benefit most when they develop an “optimal mix of distances” in knowledge exchange by concurrently exploiting external knowledge through interactions with proximate firms, and exploring new knowledge and application possibilities through interactions with distant ones (Ter Wal et al. 2016; Balland, Boschma, and Frenken 2020).

Although numerous conceptual and empirical works have discussed the path-dependent nature of knowledge sourcing behavior, little is known about the driving forces of the formation of path-breaking knowledge ties that connect cognitively or geographically distant individuals and organizations. Only recently, scholars discuss that economic geographers may go beyond the “dyadic” nature of proximity dimensions to better understand the capability of individuals and organizations in creating collaborative ties (Balland, Boschma, and Frenken 2020). This implies that creating knowledge sourcing ties between two focal distant organizations depends meaningfully on their already established knowledge ties.

Building on the notion of "the strength of weak ties" and its development in the sociology literature (Granovetter 1973; Aral and van Alstyne 2011; Aral 2016) and the recently discussed portfolio perspective (Balland, Boschma, and Frenken 2020), we argue that the driving forces of boundary-spanning knowledge ties may reside at the "triadic" (and not at the dyadic) level. This is because knowledge exchange between two cognitively distant firms requires a high degree of flexibility to deal with the complexity of communicating varied pieces of knowledge. Such relations require a "greater channel bandwidth" that can be provided by a higher degree of clustering, that is, by creating knowledge exchange ties with partners of partners (also known as strong ties) (Aral 2016). Conversely, the creation of knowledge ties between geographically distant firms may more often be driven by cognitive proximity, compensating for the lack of geographic propinquity (Balland, Boschma, and Frenken 2015; Broekel 2015). Thus, as two focal firms have already established the required interpretation scheme to communicate complementary pieces of knowledge, relations with a "smaller channel bandwidth" are beneficial for an optimal knowledge transfer, that is, knowledge ties between firms with no common partners (also known as weak ties) (Aral 2016). Last, we argue that the innovative performance of firms benefits the most if their portfolios include both strong knowledge ties with cognitively distant firms and weak ties with geographically distant ones.

To empirically examine the claims outlined above, we investigate the interaction of more than 600,000 German firms by exploring and extracting information from hyperlinks in their websites. We showed that the triadic property of relations between firms is a reliable predictor of firms' innovativeness. More precisely, strong relations bridging across different technological communities are more strongly associated with firms' innovative performance, whereas weak inter-regional relations contribute to firms' innovative capabilities. The results suggest that the joint effects of strong inter-community and weak inter-regional relations are more strongly related to firms' innovativeness compared to the combination of other knowledge relation types. The findings remain robust across multiple specifications (e.g., using patent data as an alternative proxy for innovative performance).

The structure of this paper is as follows. Section 2 reviews the notion of weak and strong ties and discusses them in the context of knowledge transfer. Section 3 highlights the relevance of relational web data reflecting inter-firm relations. Section 4 focuses on the empirical approach of the paper. Section 5 presents and discusses results. Section 6 concludes by highlighting the main results, discussing methodological contributions, and potential policy implications.

2. The strength of weak and strong knowledge ties

Evolutionary and relational approaches in economic geography have motivated an upsurge of empirical studies investigating knowledge sourcing as an "evolutionary process" (Bathelt and

Glückler 2003; Boschma and Martin 2010). This literature conceptually discusses that knowledge sourcing is path-dependent in nature (Boschma 2005; Glückler 2007) and provides empirical evidence that two firms are more likely to create a knowledge tie if they are geographically and cognitively proximate, and if they share a common collaboration partner (Giuliani 2007; Giuliani 2013; Molina-Morales et al. 2015; Lazzeretti and Capone 2016; Juhász and Lengyel 2017; Abbasiharofteh and Dyba 2018; Capone and Lazzeretti 2018; Giuliani, Balland, and Matta 2018; Abbasiharofteh and Broekel 2020). Whereas proximate firms tend to form a knowledge tie, the potential benefits drop if proximity exceeds a certain threshold (Boschma and Frenken 2010; Broekel and Boschma 2012). The notion of "proximity paradox" emphasizes that the ability to bridge geographic and cognitive distances contributes to the innovative capacity of firms by exposing them to non-redundant information (Nooteboom 2000). Although the importance of bridging geographic and cognitive distances has attracted a lot of attention in economic geography, there has been no in-depth scrutiny of whether the structural properties of bridging knowledge ties are related to the effectiveness of knowledge sourcing across space and economic activities.

The seminal work of Granovetter (1973) points toward the "strength of weak ties" in providing actors with non-redundant information. Weak ties are relations that connect parts of a given network that are not already connected. Burt (1992) builds on the strength-of-weak-ties theory and showed that actors who bridge "structural holes" are more likely to tap into diverse sources of knowledge. This is the case because occupying such network positions enables actors to monitor and control the flow of information, and consequently, gives them access to diverse information (Reagans and Zuckerman 2001; Burt 2004).

Conversely, Nelson (1989) and Krackhardt (1992; 1999) underline the importance of the "strength of strong ties." Strong ties[1] are those ties that are embedded in a clique (also known as Simmelian ties). These scholars argued that strong ties are associated with surpassing individual interests and reducing bargaining power and conflicts within groups. Thus, strong knowledge ties may contribute to the innovative performance of firms by facilitating knowledge transfer through trust building, altruistic reciprocity, and lower transaction costs (Heider 1958; Coleman 1988). This theoretical argument resonates with multiple empirical studies showing the advantages of creating strong ties (Uzzi 1997; Hansen 1999; Uzzi and Spiro 2005; Lingo and O'Mahony 2010; Obstfeld 2016).

Juxtaposing the theoretical frameworks of weak and strong ties, the former discusses the benefits of creating bridges and enjoying non-redundant information, whereas the latter

---

[1] This is important to note that some economic geography scholars argue that embeddedness in a clique represents social proximity. To be consistent with the main contributions in the literature, we use the original terminology. That is, 'weak' and 'strong' relations instead of 'socially distant' and 'socially proximate' relations.

emphasizes the normative power of groups (Krackhardt 1999). However, the main question of interest is whether this triadic property of a knowledge tie (i.e., weak and strong) facilitate knowledge sourcing behaviour that cross geographic and cognitive boundaries. To disentangle the effects of weak and strong ties in knowledge sourcing, Moody (2011) argued that besides the importance of the structural properties of a knowledge tie, one should consider the content of exchanged knowledge. He claimed that if the nature of knowledge makes it hard to transfer, actors may need to be linked by strong ties to benefit from in-group synergy effects. However, strong ties lead to knowledge redundancy and foster a lock-in situation if knowledge at hand can be easily transferred. Similarly, Aral (2016) and Aral and van Alstyne (2011) argued that the "diversity-bandwidth trade-off" may reconcile the empirical findings regarding the effectiveness of weak and strong ties in knowledge exchange. The trade-off implies that both weak ties and strong ties provide actors with access to novel information, depending on the "information environment." They suggested that actors benefit from weak ties most if an information environment entails few topics and slowly changing information. Conversely, strong ties are beneficial in an information environment with multiple topics and rapidly changing information. Bruggeman (2016) and Wu et al. (2008) extended this research by empirically showing that "simple" knowledge is better transferred through weak ties, whereas strong ties perform more efficiently when transferring rather complex knowledge. Ter Wal et al. (2016) showed the benefits of both weak specialized ties and strong diversified ties because they provides access to a diverse array of inputs. They argued that the shared interpretive schema eases the interpretation of exchanged knowledge by weak specialized ties, whereas "shared third-party ties" facilitate the interpretation of cognitively distant knowledge in closed-diverse networks. Building on these empirical findings, Aral (2016) suggested a modern strength-of-weak-ties theory that acknowledges the pioneering works (Granovetter 1973; Coleman 1988; Burt 1992) and also takes into consideration the specificity of transferred knowledge.

The extant literature on proximity suggests that geographical proximity is not the only driving force of knowledge tie formation and that one proximity dimension might compensate for the lack of the other (Boschma 2005; Balland, Boschma, and Frenken 2015). Indeed, scholars have argued that knowledge production can be seen as an optimal combination of close and distant relations (Oinas 1999; Schilling and Phelps 2007; Breschi and Lenzi 2016, 2016; Whittle, Lengyel, and Kogler 2020). For instance, Owen-Smith and Powell (2004) showed in the case of the Boston biotechnology community that firms benefit from strategic regional and inter-regional partnerships. However, it is not farfetched to assume that the threshold for optimal cognitive distance is lower for two distant firms (compared to two colocated ones) because distant firms cannot build on the synergy effects associated with geographical proximity (e.g., learning by observing and face-to-face contacts). Perhaps this explains recent empirical findings of Boschma and Balland (2020) indicating that inter-regional relations are especially useful for firms located in regions with complementary capacities. This implies that distant knowledge ties are particularly useful when firms create non-local knowledge ties that focus on a limited number of topics related to their current fields of activity. In line with Aral's (2016) argument, we expect weak knowledge ties, among cognitively proximate firms, perform more efficiently in non-local knowledge sourcing. Thus, we suggest the following hypothesis:

> *Hypothesis 1: Cognitively proximate firms that bridge geographic distance with weak knowledge ties show better innovative performance than similar ones bridging geographic distance with strong knowledge ties.*

Cognitive distance has increasingly hampered the formation of knowledge exchange ties as the complexity of knowledge increases over time (Broekel 2019). This implies that knowledge does not spill over among firms if the cognitive distance between them exceeds a certain threshold (Nooteboom 2000). Thus, firms tend to collaborate with other firms that operate in the same or in related economic activities. As a result, one might expect to observe dense knowledge ties within economic activities (hereafter, communities) and relatively fewer inter-community knowledge ties. Besides that, the path-dependent nature of knowledge sourcing increases the likelihood of intra-community tie formation (Boschma 2005; Glückler 2007; DiMaggio and Garip 2011; Tóth and Lengyel 2019). Thus, inter-community ties might need wider bridges (i.e., strong ties) to counterbalance the negative effect of cognitive distance. Explicitly, Tortoriello and Krackhardt (2010, 168) claimed that "mere bridging is not enough" and described that innovative activities need strong bridging ties to enable firms to efficiently exchange knowledge. Perhaps this is why complex economic activities and "atypical" combination of technologies increasingly concentrate in large cities (Mewes 2019; Balland et al. 2020), where creating strong ties and benefiting from the normative power of groups is more straightforward. Also, as this type of knowledge ties is more often used to exchange information on a wider range of topics across communities, one can expect that strong ties enhance the effectiveness of knowledge transfer (Aral and van Alstyne 2011; Aral 2016). Following these lines of argument, we hypothesize that:

> *Hypothesis 2: Firms that bridge cognitive distance with strong knowledge ties show better innovative performance than similar ones bridging cognitive distance with weak knowledge ties.*

Firms can create multiple strong and weak knowledge ties and they probably benefit most from knowledge sourcing strategy, whereby they form weak non-local and strong inter-community knowledge ties. Based on this rationale, we predict that:

Hypothesis 3: The combination of bridging geographic distance with weak and cognitive distance with strong knowledge ties provides the optimal configuration for the innovative performance of firms.

3. Digital footprint of inter-firm relations

The Internet "can be thought of as a self-organizing social system: individuals, with little or no central oversight, perform simple tasks: posting web pages and linking to other Web pages" (Mitchell 2009, 10). Firms post web pages to present their products, services, and costumer testimonials, but also their strategic decisions, personnel pronouncements, and relationships with other firms (Gök, Waterworth, and Shapira 2015). This textual and relational content of firm websites can then be analyzed to assess firms' innovation activity, their network relations and to map regional ecosystem (Katz and Cothey 2006; Kinne and Axenbeck 2020; Axenbeck and Breithaupt 2021; Kinne and Lenz 2021). In these analyses, hyperlinks serve as the basis to assess firms' relationships with each other (see also Vaughan and Wu 2004; Krüger et al. 2020).

Most generally, hyperlinks are considered the "basic structural element of the Internet" (Park 2003, 49). They allow users to take different paths throughout the Web, thereby revealing different communication structures among people, organizations and institutions. De Maeyer (2013) reviews an extensive set of hyperlinks studies across the fields of information science, sociology, geography, political science and media studies. In her review, one array of studies deals with uncovering the structural properties of hyperlink networks, such as their scale-free nature (Barabási and Albert 1999). The other array analyzes hyperlinks as indicators of other social phenomena by exploiting their information side-effects: "Information side-effects are by-products of data intended for one use which can be mined in order to understand some tangential, and possibly larger scale, phenomena" (Adamic and Adar 2003, 211).

One of these larger scale phenomena being explored through hyperlinks is the nature of knowledge sourcing among linked entities. Heimeriks and van den Besselaar (2006) found that "web data can be meaningful in mapping the aspects of knowledge production" on how the development of research fields is reflected in the linking patterns of scientific organizations. Vaughan et al. (2006) conducted a content analysis of 280 websites of North American IT companies and their hyperlink connections with respect to their motivation to create the hyperlinks. They found that most links indeed represent business relations, which is "strong contrast to motivations for linking to university Websites where many links are not academically related" (Vaughan, Gao, and Kipp 2006, 297). They thus concluded that hyperlinks are fertile objects for data mining in business contexts. This is in accordance with earlier studies by Vaughan and Wu (2004), who identified hyperlinks to commercial websites as a source for generating business performance indicators. Krüger et al. (2020) analyzed the hyperlink network of roughly 680,000 German firms with regard to the network position of innovative versus non-innovative firms. They also found that the majority of hyperlinks represent business relations, and that the network position of innovative firms on the Web significantly differs from the network position of non-innovative ones, when accounting for their link count, geographic distance and operationalized cognitive and organizational distances. One can therefore assume that the business relations represented through hyperlinks are a reasonable proxy for knowledge sourcing ties among firms.

However, more research needs to be done to uncover how long hyperlinks among firms generally persist and for what reasons they change. As Basole et al. (2015, 3-4) reviewed: "The social nature of the Internet has added a new data frontier in socially constructed data. […] These sources provide unprecedented access to data, updated in real time, and also include individuals' activities and interactions." For now, we do not have a clear picture, though, of how quickly these interactions change and are updated in hyperlinks. Moreover, one has to keep in mind that hyperlink system are usually comprised of organizations that are linked together around a common background, interest, or project and most likely do not reveal competitive firm relationships (Park 2003).

Still, as hyperlinks are set intentionally by firms, and creating, maintaining, or removing a hyperlink "may be viewed as acts of association, non-association or disassociation, respectively" (Rogers 2010, 117), they can be a valuable data source for analyzing knowledge sourcing relationships of firms in a business context.

4. Empirical approach

4.1. Data

In economic geography, quantitative empirical studies investigating the processes of knowledge sourcing use mostly secondary data on patents, scientific publications, and R&D projects (Bettencourt, Lobo, and Strumsky 2007; Lobo and Strumsky 2008; Strumsky and Lobo 2015; Breschi and Lenzi 2016; Juhász and Lengyel 2017; Abbasiharofteh and Broekel 2020; Balland et al. 2020). Although these studies have contributed a lot to the understanding of how firms create, maintain, and dissolve knowledge ties, more recently, scholars have called for the use of alternative firm-level databases to address unresolved research questions in economic geography (Duranton and Kerr 2018; Fritsch, Titze, and Piontek 2020). In this study, we build on the Mannheim Enterprise Panel (MUP) of 2019. The MUP is a firm panel database that covers the entire population of firms in Germany and is updated on a semi-annual basis.

In addition to firm-level characteristics, such as firm size, age, and location, the MUP includes the web addresses (URL) for 1,155,867 of 2,497,412 firms in the dataset (URL coverage of 46%). Prior analyses of this dataset (Kinne and Axenbeck 2020) showed that comparatively low URL coverage was found, for example, in the subgroup of young and small companies, whereas companies with more than 25 employees were almost completely covered. We removed firms without postal address information from our dataset and geocoded the remaining firms based on their postal codes (5-digit level) and street names (e.g., see Zandbergen 2008).

For each firm, we downloaded texts and hyperlinks contained on their respective websites. A maximum of 25 (sub-)webpages per firm website were thereby crawled and scraped. The selection of these webpages is not performed randomly but followed a simple heuristic. Preference was given to those (sub-)webpages written in German and have the shortest URL. The latter was intended to ensure that primarily more general ("top level") content is downloaded. For example, "company.com/about-us" would be downloaded before "company.com/news/2019/august." After excluding erroneous downloads and potentially misleading redirects (see Kinne and Axenbeck 2020) from the data, 633,523 firms remained in the dataset.

We then constructed a directed network of firms based on the hyperlink references found on their websites (3,062,670 ties). As we used these hyperlink connections as a proxy for the presence of knowledge sourcing ties, we took a more conservative approach and included only reciprocated ties in the network (1,363,305 ties, whereas the number of firms remains constant: 633,523). This means that a tie exists between firm A and firm B only if both firms include hyperlinks to each other's website on their own website. This approach is very much in line with the rationale behind weak and strong ties. Two firms are strongly tied (Simmelian tied) to each other if they are reciprocally connected and reciprocally connected to at least one common firm (Krackhardt 1999). Accordingly, two firms are weakly tied (non-Simmelian tied) if they are reciprocally connected and are not reciprocally connected to one or several common firms. Figure 1 maps the locations of all firms in our dataset (high density areas shown in brighter colors) and a random sample of hyperlink connections between them. For the purpose of clarity, the hyperlinks in Figure 1 were aggregated using a graph bundling method based on kernel density estimation (Hurter, Ersoy, and Telea 2012).

Figure 1. Visualization of firm locations and the intensity of inter-firm hyperlink connections at the national level.

4.2. Dependent variable

As an alternative innovation indicator at the company level, we adopted the InnoProb approach described by Kinne and Lenz (2019) and shown schematically in Figure 2. It is a web-scraping based method in which the text content of the websites of the companies being investigated is downloaded and then assessed by an artificial neural network. The artificial neural network acts as a text classification model, which analyzes the input texts and then outputs a “predicted product innovator probability.” More precisely, it estimates the likelihood of the examined texts’ originating from a company that has launched new or significantly improved products to the market. This ability was previously “taught” to the model in a training phase in which the texts of companies that participated in a traditional innovation survey serve as training data. To this end, the German Community Innovation Survey (CIS), a questionnaire-based survey in the style of the Oslo Manual (OECD 2018), was used. In the survey, about 12,000 companies were asked about their innovation activities, including whether they are so-called product innovators as described above.

For the InnoProb model, the websites of the surveyed companies are downloaded, the found texts are vectorized according to the tf-idf scheme (e.g., see Manning, Raghavan P, and Schutze 2009) and used along with the information on whether the company is a product innovator as training data for a deep neural network (Kinne and Lenz 2019). During the training phase, the model “learns” which words and word combinations characterize a product innovator. After training, the model can assess the website texts of any (out-of-sample) firm and predict its "product innovator probability." These "InnoProb scores" range from 0.0 (unlikely product innovator) to 1.0 (likely product innovator) and have been tested for their robustness and validity against several traditional innovation indicators ranging from patent data to official statistics (see Kinne and Lenz 2019).

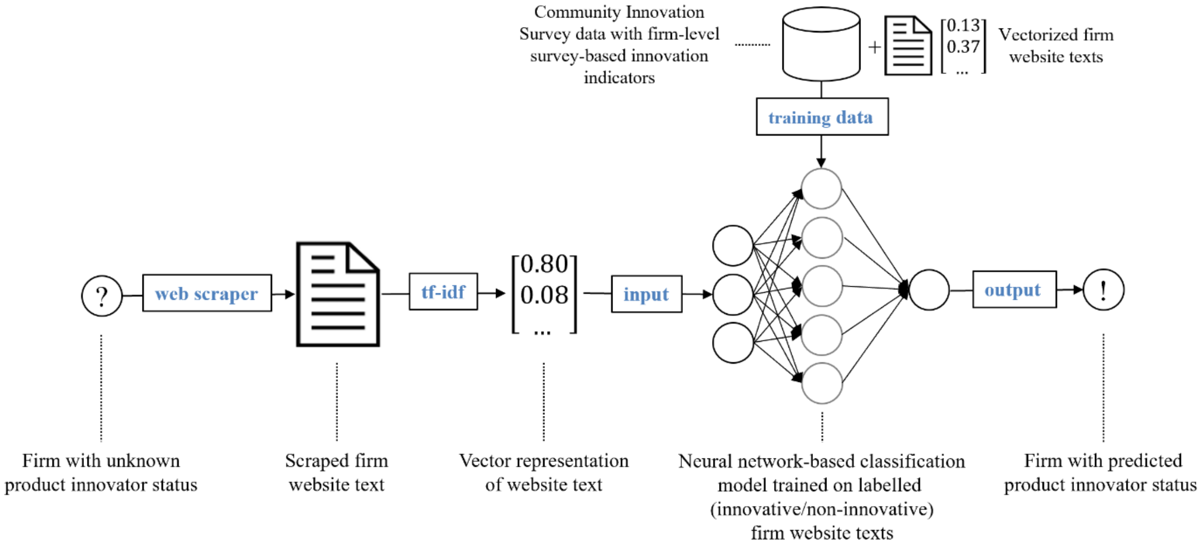

Figure 2. Schematic representation of the InnoProb model for evaluating text from corporate websites. Adapted from Kinne and Lenz (2019).

For this study, we calculated the firm-level InnoProb scores for all 633,523 companies in our dataset (Figure 3). Subsequently, we binarized these raw InnoProb scores using a classification threshold that corresponds to the 90th percentile (0.5) to obtain a binary (dummy) dependent variable (*INNOVATIVE*). Compared to the recommended classification threshold of 0.4 by Kinne and Lenz (2019), we opted for a rather strict classification approach to minimize the share of false positives in the group of firms that we classify as being innovative. As will be discussed, we selected multiple thresholds for the InnoProb classification and replicated the empirical analysis to ensure the robustness of our empirical results.

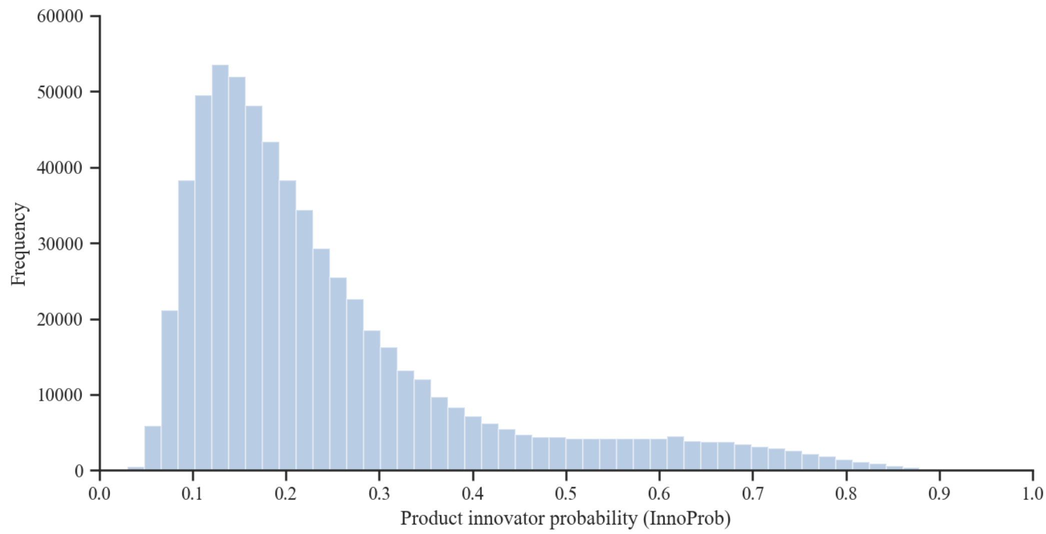

Figure 3. Distribution of product innovator probability (InnoProb) scores.

4.3. Independent variables

The main variables of interest approximate the extent to which each firm is connected to others through weak and strong bridges across geographic and cognitive distances. We defined geographic bridging ties as those inter-firm hyperlink relations that cross regional boundaries (NUTS-2 level). It is, however, less straightforward to identify ties that bridge cognitive distances. We argue that firms are more likely to create a knowledge exchange tie when they operate in the same or related economic activities. The importance of cognitive proximity as a driving force of collaborative ties formation has been theoretically discussed (Boschma 2005) and empirically shown by multiple studies (Ter Wal 2014; Balland, Belso-Martínez, and Morrison 2015; Lazzeretti and Capone 2016; Juhász and Lengyel 2017). This translates into a greater number of ties between cognitively proximate firms, and a smaller number of ties between cognitively distant ones.

A community detection algorithm[2] enabled us to identify densely connected communities. Building on this, we define ties that connect different communities as the ones that bridge cognitive distances. Following Krackhardt (Krackhardt 1999), Figure 4 provides a stylized representation of the four bridging ties: weak inter-regional (*WIR*), strong inter-regional (*SIR*), weak inter-community (*WIC*), and strong inter-community (*SIC*) bridging ties. To measure the

[2] We used the multilevel community detection algorithm because it is efficient (time complexity: $\mathcal{O}(N\,logN)$) and provides the most reliable results given the size of the investigated network (Yang, Algesheimer, and Tessone 2016).

extent to which each firm weakly or strongly bridges geographic and cognitive distances, we use the *E-I index* suggested by Krackhardt and Stern (1988).

$$E - I\ index\ = \frac{E_i\ -\ I_i}{E_i\ +\ I_i} \tag{1}$$

E denotes the number of ties that firm *i* has with firms that belong to groups other than firm *i*'s group (i.e., other regions or technological communities). Accordingly, *I* represents the number of ties that *i* has with other firms belonging to the same group (i.e., same region or technological community). The minimum value corresponds to minus one if a firm has only within-group ties, and the maximum estimated value corresponds to one if a firm has only between-group ties. We estimated the bridging index for each type of bridging ties separately. After estimating *SIR*, *WIR*, *SIC,* and *WIC,* we standardized the four variables by expressing them as z-scores[3] (mean:0 and standard deviation:1). This enabled us to better interpret and compare the coefficients of these four variables estimated by regression models.

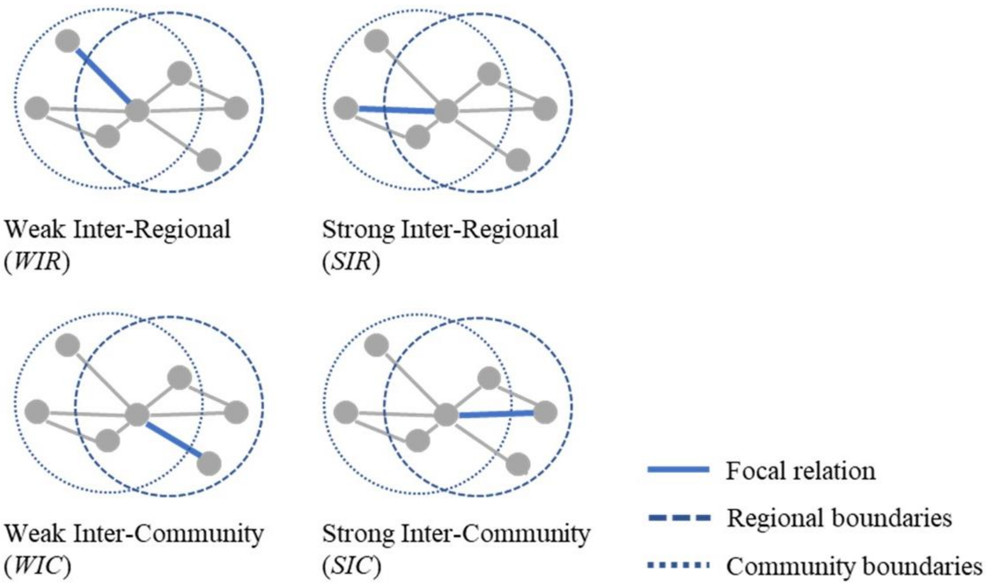

Figure 4. Schematic illustration of the four bridging types.

4.4. Controls

To ensure the robustness of our results, we constructed multiple control variables. Following similar empirical studies, the variables could be categorized into three groups: 1) network[4], 2) individual, and 3) regional levels (Lobo and Strumsky 2008; Tortoriello and Krackhardt 2010;

[3] z-score = (x-x̄)/sd(x) where x̄ and sd(x) are the mean and standard deviation of x, respectively.
[4] For estimating network level variables, we used the 'igraph' R package (Csardi and Nepusz 2006).

Breschi and Lenzi 2016; Balland et al. 2018; Bergé, Carayol, and Roux 2018; van der Wouden and Rigby 2019; Abbasiharofteh and Broekel 2020). Table 1 provides a description of these control variables (Appendix A provides descriptive statistics and the associated correlation matrix).

Category	Type (range)	Method	Variable name
Network level	Continuous (0.69- 11.85)	The log-transformed count of relations (reciprocal hyperlinks) a firm has; also known as degree centrality.	*DEGREE*
	Continuous (0.00-18.95)	The log-transformed sum of *DEGREE* of a firm's neighbors.	*ALTER*
	Continuous (0.00-1.00)	The probability that the neighbors of a firm are connected (Wasserman and Faust 1994).	*TRANSITIVITY*
Individual level	Continuous (0.65-6.93)	Age* of firm in years (log-transformed).	*AGE*
Regional level	Dummy (0 or 1)	This variable takes the value of 1 if a firm is located in a metropolitan region (defined by Eurostat) and takes the value of 0 otherwise.	*METROPOL*
	Dummy (0 or 1)	This variable takes the value of 1 if a firm is located in the former East Germany (GDR) and takes the value of 0 otherwise.	*EAST*
	Continuous (0.00-8.28)	Log-transformed count of other firms within a 1 km radius around a given firm (all firms in the MUP dataset are considered, not only the ones in our network).	*FDENSITY*
	Continuous (8.41-10.7)	The log-transformed number of firms in the NUTS-2 region, in which a given firm is located.	*RFDENSITY*

Table 1. Control variables.

* A small number of companies have very large *AGE* values, which represent old family businesses and breweries.

5. Results and discussion

We ran multiple logit models to investigate whether the capability of firms to bridge geographic and cognitive distances and the characteristics of bridging ties statistically correlate with their innovation performance. The results are shown in Table 2. The goodness of fit improves as more variables are introduced and the full model provides the best goodness of fit.

Regarding the network level controls, it seems that the number of inter-firm relations is positively related to a firm's probability of being a product innovator. This finding resonates with the scholarly debate about how firms occupying central positions are exposed to a wider range of sources of information that can be used in the innovation process (Bergé, Carayol, and Roux 2018). However, the sign and the significance of coefficients capturing the effects of clustering and the relations of a firm's neighborhood (i.e., the sum of adjacent nodes' relations) change across the model specifications in Table 2. Thus, we refrain from interpreting and discussing the coefficients of these two variables.

The reported coefficients of the individual and regional control variables are mostly in line with our expectations and similar to those of other empirical studies. The age of firms is negatively related to the capability of firms to innovate. This finding is congruent with the empirical evidence from the scaling literature claiming that young firms grow and innovate at a higher pace, relative to the later stages of their development (Bettencourt et al. 2007; West 2017).

Moreover, we found that the two variables capturing firm density (*METROPOL* and *FDENSITY*) are positively associated with the product innovator probability of firms. The extensive body of literature on agglomeration economies provides evidence that colocated firms benefit from a wide range of externalities, such as knowledge spillovers, input and output linkages, and skilled labor pooling (Duranton and Puga 2004). Finally, our results suggest that firms in the Eastern part of Germany are less innovative, compared to their Western counterparts, which may reflect cultural, sectoral, and socio-economic differences, as well as the embeddedness of firms in various innovation systems, with that of East Germany being more centralized (van Hoorn and Maseland 2010; Fritsch and Slavtchev 2011; Wenau, Grigoriev, and Shkolnikov 2019; Abbasiharofteh and Broekel 2020). Taken together, the estimated coefficients of our control variables reflect our expectations derived from previous studies and theoretical considerations. We are therefore confident that our empirical setting can be used to confirm or reject the formulated hypotheses of this study.

We introduced the variables of interest in our models in a stepwise manner. Since the results are consistent across all models and given that the full model (Model 6) provides the best goodness of fit, we interpret and discuss the results of the full model. The four variables that capture the effect of bridging with different attributes are positively correlated with the product innovator probability of firms, whereas the positive coefficient of weak inter-community ties is not statistically significant. We are particularly interested in investigating the magnitude of effects across different bridging types. Our results suggest that the share of weak inter-regional bridging ties is more strongly correlated with the dependent variable than the one of strong inter-regional ties. This confirms Hypothesis 1. Since mutually tied firms located in different regions cannot rely on advantages associated with geographical proximity, they may compensate for this disadvantage by interacting with firms that have the same or similar portfolios (Balland, Boschma, and Frenken 2015; Boschma and Balland 2020). Thus, weak inter-regional ties may favor innovative behavior because inter-regional ties are channels for exchanging knowledge between cognitively proximate firms.

Furthermore, the reported coefficient of *SIC* seems to be significantly related to *INNOVATIVE* whereas the one of *WIC* are not significantly related. This confirms Hypothesis 2. Whereas most economic geography studies focus on individual and regional factors to account for the innovative capability of firms and regions, the relevance of the meso-level (community level) has often been, to some extent, ignored. A recent economic geography study discussed the importance of this level of analysis and showed that collaborative ties among specialized and cognitively distant communities are a critical factor enabling regions to introduce unconventional innovations (Abbasiharofteh, Kogler, and Lengyel 2020). However, this study said nothing about which attributes of such inter-community bridging ties may improve the quality of knowledge sourcing. Our results suggest that firms benefit from inter-community relations only if such ties are strong bridging ones (i.e., shared with at least a common third). As discussed earlier, since ties that bridge communities might be associated with bridging cognitively distant firms, this may increase uncertainty, the complexity of interaction, and the number of included topics. This finding resonates with the ones of Aral and van Alstyne (2011) who argued that strong ties ("greater channel bandwidth", in their language) facilitate knowledge sourcing in such settings.

	Dependent variable: INNOVATIVE					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
SIR		0.1839***				0.1606***
		(0.0061)				(0.0062)
WIR			0.2045***			0.1901***
			(0.0069)			(0.0070)
SIC				0.0876***		0.0813***
				(0.0058)		(0.0059)
WIC					-0.0031	0.0039
					(0.0060)	(0.0061)
DEGREE	0.3605***	0.2742***	0.4124***	0.3702***	0.3618***	0.3395***
	(0.0073)	(0.0079)	(0.0075)	(0.0074)	(0.0077)	(0.0086)
ALTER	0.0057***	0.0077***	-0.0115***	0.0053***	0.0057***	-0.0090***
	(0.0012)	(0.0012)	(0.0013)	(0.0012)	(0.0012)	(0.0013)
TRANSITIVITY	0.0739***	-0.1871***	0.0965***	0.2280***	0.0776***	0.0061
	(0.0277)	(0.0298)	(0.0278)	(0.0295)	(0.0286)	(0.0326)
AGE	-0.3621***	-0.3567***	-0.3566***	-0.3600***	-0.3621***	-0.3500***
	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0069)
FDENSITY	0.1586***	0.1636***	0.1611***	0.1586***	0.1586***	0.1652***
	(0.0040)	(0.0041)	(0.0041)	(0.0040)	(0.0040)	(0.0041)
RFDENSITY	0.1888***	0.2068***	0.2110***	0.1909***	0.1888***	0.2269***
	(0.0105)	(0.0106)	(0.0106)	(0.0105)	(0.0105)	(0.0106)
METROPOL	0.1913***	0.1846***	0.1844***	0.1908***	0.1913***	0.1789***
	(0.0148)	(0.0148)	(0.0148)	(0.0148)	(0.0148)	(0.0148)
EAST	-0.1695***	-0.1550***	-0.1675***	-0.1659***	-0.1695***	-0.1523***
	(0.0153)	(0.0154)	(0.0154)	(0.0154)	(0.0153)	(0.0154)
Constant	-5.9594***	-6.0129***	-6.1531***	-6.0160***	-5.9612***	-6.2329***
	(0.1676)	(0.1678)	(0.1679)	(0.1677)	(0.1676)	(0.1682)
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-114,082.6000	-113,609.9000	-113,630.4000	-113,970.5000	-114,082.5000	-113,150.9000
Akaike Inf. Crit.	228,265.2000	227,321.7000	227,362.8000	228,043.0000	228,266.9000	226,409.8000

*p<0.1; **p<0.05; ***p<0.01

Table 2. Logit regression estimation results.

We also investigated the joint effects of different types of bridging ties on the product innovator probabilities of firms. Given that interpreting the joint effects of dummy variables is more straightforward, we followed a common practice (Cattani and Ferriani 2008; Juhász, Tóth, and Lengyel 2020) and binarized our four variables capturing bridging effects. More specifically, these transformed variables take the value of one if the value for the respective bridging type is greater than the 90th percentile of the original variable (otherwise the value is zero).

Table 3 provides the estimation results for six different logit regression specifications covering all possible dyadic interaction terms between bridging variables. The reported coefficients of our control variables are similar in sign and significance to the ones presented above. Thus, for the sake of brevity we only report the coefficients capturing the effects of bridging variables and their interactions.

The results confirm Hypothesis 3, given that the joint effects of strong inter-community and weak inter-regional bridging ties are positively related to the innovativeness of firms. More interestingly, the joint effects of strong inter-regional and weak inter-community bridging ties are negatively correlated, meaning these two bridging types substitute the effect of each other, whereas strong inter-community and weak inter-regional ones complement the effects of each other[5]. Although the theoretical argument and the empirical findings of economic geography studies point toward the importance of both relations with firms in other regions and with colocated ones (Oinas 1999; Nooteboom 2000; Bathelt, Malmberg, and Maskell 2004; Boschma 2005; Bathelt and Turi 2011; Breschi and Lenzi 2013), our results contribute to this scholarly debate by suggesting that, while firms have to bridge regional and cognitive distances to be innovative, the structural properties of bridges may indeed determine the effectiveness of such knowledge sourcing efforts[6].

[5] It is important to note that we also included all possible triadic and quadratic interaction terms (see appendices), and this finding remained robust across all these specifications.
[6] The joint effects of strong inter-regional and weak inter-regional bridging ties are also positively correlated with the dependent variable. However, the coefficient of this variable loses significance in an extended interaction model (see Appendix E).

	Dependent variable: INNOVATIVE					
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
SIR (dummy)	0.3130***	0.2548***	0.4127***			
	(0.0217)	(0.0212)	(0.0311)			
SIC (dummy)	0.1574***			0.0882***	0.1702***	
	(0.0200)			(0.0195)	(0.0269)	
WIR (dummy)		0.3251***		0.3105***		0.3518***
		(0.0145)		(0.0146)		(0.0169)
WIC (dummy)			-0.0134		-0.0263**	0.0270
			(0.0128)		(0.0132)	(0.0166)
SIR×SIC	-0.0372					
	(0.0395)					
SIR×WIR		0.1955***				
		(0.0390)				
SIR×WIC			-0.1341***			
			(0.0373)			
SIC×WIR				0.4159***		
				(0.0378)		
SIC×WIC					0.0168	
					(0.0339)	
WIR×WIC						0.0040
						(0.0253)
Constant	-6.0857***	-6.4398***	-6.0784***	-6.2920***	-5.9547***	-6.3630***
	(0.1679)	(0.1688)	(0.1679)	(0.1686)	(0.1677)	(0.1688)
Controls	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-113,894.5000	-113,624.2000	-113,923.1000	-113,656.4000	-114,025.8000	-113,773.2000
Akaike Inf. Crit.	227,895.0000	227,354.5000	227,952.2000	227,418.7000	228,157.5000	227,652.4000

*p<0.1; **p<0.05; ***p<0.01

Table 3. Logit regression estimation results with interaction terms.

5.1. Robustness check

In the original models (Table 2), we constructed the dependent variable based on the 90th percentile of the InnoProb web-based innovation indicator. To ensure the reliability of our results, we carried out a first robustness check by regressing a similar set of independent and control variables against dependent variables constructed using alternative thresholds (i.e., 50th and 75th percentiles). Interestingly, while the positive correlation of inter-regional bridging ties and *INNOVATIVE* seems robust against different thresholds, the positive effects of strong inter-community bridging diminish when we lower the threshold to the median value of InnoProb (see Appendix B). Moreover, the variable capturing the effect of weak inter-community bridging ties is even negatively correlated with the dependent variable when we lower the classification threshold of *INNOVATIVE*. When changing the threshold value to the 75th percentile of InnoProb, the effect of strong inter-community bridges becomes positive and statistically significant, whereas the findings regarding the effect of weak and strong inter-regional ties are consistent with the original model. Thus, we conjecture that relations connecting firms from two different communities may be positively related with exceptionally innovative performances. This result resonates with a finding from of Abbasiharofteh et al. (2020) who showed that inter-community collaborations enable inventors to introduce more atypical inventions. The models with interaction terms show that the joint effects of weak inter-regional and strong inter-communities bridging ties remain positive and statistically significant.

As estimating firms' innovation activities using machine-learning techniques is not a common practice in economic geography, we also used a more established indicator of innovative capabilities as a second robustness test for our findings. We retrieved patent data from the European Patent Office (EPO) and, due to the declining economic and technological value of aging patents, limited the dataset to patents with grant dates between 2006 and 2017 (a 10-year time period corresponding to the average lifetime of patents in our database). We then determined the "patent holder status" of each company and used it as an alternative dependent variable. That is, the dependent variable takes the value of 1 if a company holds at least one patent and 0 otherwise. The results of this robustness check (see Appendix C) are consistent with the results of our original models, while the difference between the magnitudes of effects decreases. We are therefore confident that our initial results were not driven by the method we used to approximate the innovativeness of firms.

The innovative performance of firms may arguably be positively influenced by those bridging ties that simultaneously cross geographic and cognitive distances, because inter-community and inter-regional bridging ties are not mutually exclusive. In a third robustness test, we controlled for this overlap by excluding all ties that bridge both communities and regions and repeated logit regressions. The findings (see Appendix D) remain robust in this specification.

Last, we added alternative variables at the network and regional levels. For instance, we constructed a variable based on Burt's constraint (2004) for each firm. Burt's constraint increases if a firm has more redundant relations. We created a variable corresponding to the size of the network component (the number of firms) in which a firm is located. In addition, to control for the specialization of firms' regions, we estimated the average value of regional technologies' related density (Boschma, Balland, and Kogler 2015). Technologies are proxied by patent technological codes (the 4-digit level) filed in each region in the past 10 years (van der Wouden and Rigby 2019)[7]. Introducing these extra network and regional level variables did not challenge our main findings regarding the relation between bridging ties and their interactions, and the innovative performance of companies.

6. Conclusion

The proximity framework has provided a conceptual engine to analyze and understand the driving forces of knowledge tie formation (Boschma 2005; Balland, Boschma, and Frenken 2020). Whereas it has been discussed that boundary-spanning knowledge ties may maximize the likelihood of innovative performance, to the best of our knowledge, a necessary structural configuration for this type of knowledge sourcing has not been empirically addressed. To this end, this paper contributes to the ongoing scholarly discourse by arguing conceptually and showing empirically that the interplay between the nature of exchanged knowledge and the triadic property of dyadic knowledge relations may play a role in increasing the capability of firms to bridge cognitive and geographic distances. Our results resonate with Granovetter's (1973) argument that "treating only the strength of ties ignores, for instance, all the important issues involving their content" (p. 1378). More precisely, our results show that knowledge exchange between cognitively distant firms may be more strongly associated with the diversity of topics and knowledge communication complexity. Therefore, strong knowledge ties seem to provide the most optimal configuration, perhaps by increasing the capacity of monitoring, communicating, and interpreting (Krackhardt 1999; Ter Wal et al. 2016). Conversely, weak knowledge ties seem to be more efficient at bridging geographic distance, perhaps because such relations are among firms operating in the same or related fields with common interpretation schemes. In addition, firms' innovative performance is positively related to the joint effects of the two types of knowledge relations.

Methodologically, this paper integrates techniques developed in the machine learning community to create a proxy for innovative performance. Whereas multiple disciplines such as computational social science, networks science, and applied economics have started to benefit from machine learning techniques (Muscoloni et al. 2017; Emmert-Streib et al. 2020; Storm, Baylis, and Heckelei 2020), it seems that economic geographers have, to some extent, overlooked the power of such techniques to mine and analyze much needed micro-level data (Duranton and Kerr 2018; Fritsch, Titze, and Piontek 2020). Economic geographers and

[7] The results are available upon request.

regional studies scholars can take this study as a point of departure to enhance the diversity of available data and the methodological toolbox.

Having reviewed the main contributions of the paper, this study is not free of limitations. We used a mutual inter-firm's hyperlink network as a proxy for knowledge exchange. Although this assumption does not seem too farfetched, future research needs to study this issue by mapping various types of inter-firm relations (e.g., social, economic, knowledge relations) and by investigating their impact on the formation of inter-firm hyperlinks. Moreover, the developed Innovator Probability Index that proxies the innovative capabilities of firms is agnostic about process innovation which could be considered for the further development of Innovator Probability Index. Another limitation of this paper concerns the static nature of the network data. Networks are dynamic phenomena, and they may change over time. Although we do not expect to see a fundamental change in the inter-firm hyperlink networks in a short time, we do, however, suggest replicated testing of similar data retrieved in another time-window. Having relational data in two consequent time-windows also provides an abundance of possibilities for longitudinal studies. Finally, we encourage economic geographers to use different datasets (e.g., patent and R&D data) to empirically test the external validity of our results.

While the importance of collaboration among cognitively distant actors at the local level has been proven (Singh 2005), our study suggests that such knowledge sourcing behaviors are more likely to succeed if they are embedded in strong knowledge relations. The relevance of the property of the triadic-level of knowledge relations may be deployed in the context of place-based and mission-oriented innovation policies. These policies encourage collaborations among diverse stakeholders to discover unexploited potentials and to find solutions for grant societal challenges (Foray 2018; Mazzucato 2018; Hekkert et al. 2020). An improved understanding of the interaction between strong and weak ties, as well as the nature of collaboration, may help policymakers overcome the inertia caused by cognitive and geographic distances that challenge the involvement of diverse stakeholders in joint projects.

References

Abbasiharofteh, M., and Broekel, T. 2020. Still in the shadow of the wall? The case of the Berlin biotechnology cluster. *Environment and Planning A: Economy and Space* 46 (3). doi:10.1177/0308518X20933904.

Abbasiharofteh, M., and Dyba, W. 2018. Structure and significance of knowledge networks in two low-tech clusters in Poland. *Regional Studies, Regional Science* 5 (1): 108–16. doi:10.1080/21681376.2018.1442244.

Abbasiharofteh, M., Kogler, D. F., and Lengyel, B. 2020. Atypical Combination of Technologies in Regional Co-inventor Networks. *Papers in Evolutionary Economic Geography (PEEG)* 20.55. http://econ.geo.uu.nl/peeg/peeg2055.pdf.

Adamic, L. A., and Adar, E. 2003. Friends and neighbors on the Web. *Social Networks* 25 (3): 211–30. doi:10.1016/S0378-8733(03)00009-1.

Aral, S. 2016. The Future of Weak Ties. *American Journal of Sociology* 121 (6): 1931–39. doi:10.1086/686293.

Aral, S., and van Alstyne, M. 2011. The Diversity-Bandwidth Trade-off. *American Journal of Sociology* 117 (1): 90–171. doi:10.1086/661238.

Axenbeck, J., and Breithaupt, P. 2021. Innovation indicators based on firm websites-Which website characteristics predict firm-level innovation activity? *PloS one* 16 (4): e0249583. doi:10.1371/journal.pone.0249583.

Balland, P.-A., Belso-Martínez, J. A., and Morrison, A. 2015. The Dynamics of Technical and Business Knowledge Networks in Industrial Clusters: Embeddedness, Status, or Proximity? *Economic Geography* 92 (1): 35–60. 10.1080/00130095.2015.1094370.

Balland, P.-A., Boschma, R., Crespo, J., and Rigby, D. L. 2018. Smart specialization policy in the European Union: Relatedness, knowledge complexity and regional diversification. *Regional Studies* 35 (4): 1–17. doi:10.1080/00343404.2018.1437900.

Balland, P.-A., Boschma, R., and Frenken, K. 2015. Proximity and Innovation: From Statics to Dynamics. *Regional Studies* 49 (6): 907–20. doi:10.1080/00343404.2014.883598.

Balland, P.-A., Boschma, R., and Frenken, K. 2020. *Proximity, Innovation and Networks: A Concise Review and Some Next Steps.* Utrecht: Department of Human Geography and Spatial Planning, Group Economic Geography, Utrecht University.

Balland, P.-A., Jara-Figueroa, C., Petralia, S. G., Steijn, M. P. A., Rigby, D. L., and Hidalgo, C. A. 2020. Complex economic activities concentrate in large cities. *Nature human behaviour.* doi:10.1038/s41562-019-0803-3.

Barabási, A.-L., and Albert, R. 1999. Emergence of Scaling in Random Networks. *Science* 286 (5439): 509–12. doi:10.1126/science.286.5439.509.

Basole, R. C., Russell, M. G., Huhtamäki, J., Rubens, N., Still, K., and Park, H. 2015. Understanding Business Ecosystem Dynamics. *ACM Transactions on Management Information Systems* 6 (2): 1–32. doi:10.1145/2724730.

Bathelt, H., and Glückler, J. 2003. Toward a relational economic geography. *Journal of Economic Geography* 3 (2): 117–44. doi:10.1093/jeg/3.2.117.

Bathelt, H., Malmberg, A., and Maskell, P. 2004. Clusters and knowledge: Local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography* 28 (1): 31–56. doi:10.1191/0309132504ph469oa.

Bathelt, H., and Turi, P. 2011. Local, global and virtual buzz: The importance of face-to-face contact in economic interaction and possibilities to go beyond. *Geoforum* 42 (5): 520–29. doi:10.1016/j.geoforum.2011.04.007.

Bergé, L., Carayol, N., and Roux, P. 2018. How do inventor networks affect urban invention? *Regional Science and Urban Economics* 71:137–62.

Bettencourt, L. M. A., Lobo, J., Helbing, D., Kühnert, C., and West, G. B. 2007. Growth, innovation, scaling, and the pace of life in cities. *Proceedings of the National Academy of Sciences of the United States of America* 104 (17): 7301–6. doi:10.1073/pnas.0610172104.

Bettencourt, L. M.A., Lobo, J., and Strumsky, D. 2007. Invention in the city: Increasing returns to patenting as a scaling function of metropolitan size. *Research Policy* 36 (1): 107–20. doi:10.1016/j.respol.2006.09.026.

Boschma, R. 2005. Proximity and Innovation: A Critical Assessment. *Regional Studies* 39 (1): 61–74. doi:10.1080/0034340052000320887.

Boschma, R., and Balland, P.-A. 2020. *Complementary Inter-Regional Linkages and Smart Specialization: an Empirical Study on European Regions.* Utrecht: Department of Human Geography and Spatial Planning, Group Economic Geography, Utrecht University.

Boschma, R., and Frenken, K. 2010. The spatial evolution of innovation networks. A proximity perspective. In *The Handbook of Evolutionary Economic Geography*, ed. R. Boschma and R. Martin, 120–35: Edward Elgar Publishing.

Boschma, R., and Martin, R., eds. 2010. *The handbook of evolutionary economic geography.* Cheltenham, U.K: Edward Elgar Pub. http://site.ebrary.com/lib/alltitles/docDetail.action?docID=10415747.

Breschi, S., and Lenzi, C. 2013. Local Buzz Versus Global Pipelines and the Inventive Productivity of US Cities. In *The Geography of Networks and R&D Collaborations*, ed. T. Scherngell, 299–315. Advances in Spatial Science. Cham: Springer International Publishing.

———. 2016. Co-invention networks and inventive productivity in US cities. *Journal of Urban Economics* 92:66–75.

Breschi, S., and Lissoni, F. 2009. Mobility of skilled workers and co-invention networks: An anatomy of localized knowledge flows. *Journal of Economic Geography* 9 (4): 439–68. doi:10.1093/jeg/lbp008.

Broekel, T. 2015. The Co-evolution of Proximities - A Network Level Study. *Regional Studies* 49 (6): 921–35. doi:10.1080/00343404.2014.1001732.

———. 2019. Using structural diversity to measure the complexity of technologies. *PloS one* 14 (5): e0216856. doi:10.1371/journal.pone.0216856.

Broekel, T., and Boschma, R. 2012. Knowledge networks in the Dutch aviation industry: The proximity paradox. *Journal of Economic Geography* 12 (2): 409–33. doi:10.1093/jeg/lbr010.

Bruggeman, J. 2016. The Strength of Varying Tie Strength: Comment on Aral and Van Alstyne. *American Journal of Sociology* 121 (6): 1919–30. doi:10.1086/686267.

Burt, R. S. 1992. *Structural holes: The social structure of competition.* Cambridge, Mass. Harvard Univ. Press.

———. 2004. Structural Holes and Good Ideas. *American Journal of Sociology* 110 (2): 349–99. doi:10.1086/421787.

Capone, F., and Lazzeretti, L. 2018. The different roles of proximity in multiple informal network relationships: Evidence from the cluster of high technology applied to cultural goods in Tuscany. *Industry & Innovation* 25 (9): 897–917. doi:10.1080/13662716.2018.1442713.

Cattani, G., and Ferriani, S. 2008. A Core/Periphery Perspective on Individual Creative Performance: Social Networks and Cinematic Achievements in the Hollywood Film Industry. *Organization Science* 19 (6): 824–44. doi:10.1287/orsc.1070.0350.

Clark, G. L., Feldman, M. P., Gertler, M. S., and Wójcik, D., eds. 2018. *The New Oxford Handbook of Economic Geography.* Oxford: Oxford University Press.

Coleman, J. S. 1988. Social capital in the creation of human capital. *American Journal of Sociology* 94: 95–120.

Csardi, G., and Nepusz, T. 2006. The igraph software package for complex network research: InterJournal, Complex Systems 1695. http://igraph.org.

deMaeyer, J. 2013. Towards a hyperlinked society: A critical review of link studies. *New Media & Society* 15 (5): 737–51. doi:10.1177/1461444812462851.

DiMaggio, P., and Garip, F. 2011. How Network Externalities Can Exacerbate Intergroup Inequality. *American Journal of Sociology* 116 (6): 1887–1933. doi:10.1086/659653.

Duranton, G., and Kerr, W. 2018. The Logic of Agglomeration. In *The new Oxford handbook of economic geography*, ed. G. L. Clark, M. P. Feldman, M. S. Gertler, and D. Wójcik. First edition, 347–65. Oxford: Oxford University Press.

Duranton, G., and Puga, D. 2004. Chapter 48 Micro-foundations of urban agglomeration economies. In *Cities and Geography*. Vol. 4, 2063–2117. Handbook of Regional and Urban Economics: Elsevier.

Emmert-Streib, F., Yang, Z., Feng, H., Tripathi, S., and Dehmer, M. 2020. An Introductory Review of Deep Learning for Prediction Models With Big Data. *Frontiers in Artificial Intelligence* 3:831.

Foray, D. 2018. Smart specialization strategies as a case of mission-oriented policy—a case study on the emergence of new policy practices. *Industrial and Corporate Change* 27 (5): 817–32. doi:10.1093/icc/dty030.

Fritsch, M., and Slavtchev, V. 2011. Determinants of the Efficiency of Regional Innovation Systems. *Regional Studies* 45 (7): 905–18. doi:10.1080/00343400802251494.

Fritsch, M., Titze, M., and Piontek, M. 2020. Identifying cooperation for innovation—a comparison of data sources. *Industry & Innovation* 27 (6): 630–59. doi:10.1080/13662716.2019.1650253.

Giuliani, E. 2007. The selective nature of knowledge networks in clusters: Evidence from the wine industry. *Journal of Economic Geography* 7 (2): 139–68. doi:10.1093/jeg/lbl014.

Giuliani, E. 2013. Network dynamics in regional clusters: Evidence from Chile. *Research Policy* 42 (8): 1406–19. doi:10.1016/j.respol.2013.04.002.

Giuliani, E., Balland, P.-A., and Matta, A. 2018. Straining but not thriving: Understanding network dynamics in underperforming industrial clusters. *Journal of Economic Geography* 30:147–72.

Glückler, J. 2007. Economic geography and the evolution of networks Johannes. *Journal of Economic Geography* 7 (5): 619–34. https://doi.org/10.1093/jeg/lbm023.

Gök, A., Waterworth, A., and Shapira, P. 2015. Use of web mining in studying innovation. *Scientometrics* 102:653–71.

Granovetter, M. 1973. The strength of weak ties. *American Journal of Sociology* 78: 1360–80.

Hansen, M. T. 1999. The Search-Transfer Problem: The Role of Weak Ties in Sharing Knowledge across Organization Subunits. *Administrative Science Quarterly* 44 (1): 82. doi:10.2307/2667032.

Heider, F. 1958. *The psychology of interpersonal relations.* New Jersey: Lawrence Erblaum Associates, Publishers.

Heimeriks, G., and van den Besselaar, P. 2006. Analyzing hyperlinks networks: The meaning of hyperlink based indicators of knowledge production. *Cybermetrics* 10.

Hekkert, M. P., Janssen, M. J., Wesseling, J. H., and Negro, S. O. 2020. Mission-oriented innovation systems. *Environmental Innovation and Societal Transitions* 34:76–79.

Hurter, C., Ersoy, O., and Telea, A. 2012. Graph Bundling by Kernel Density Estimation. *Computer Graphics Forum* 31 (3pt1): 865–74. doi:10.1111/j.1467-8659.2012.03079.x.

Juhász, S., and Lengyel, B. 2017. Creation and persistence of ties in cluster knowledge networks. *Journal of Economic Geography* 121: 1203–26. 10.1093/jeg/lbx039.

Juhász, S., Tóth, G., and Lengyel, B. 2020. Brokering the core and the periphery: Creative success and collaboration networks in the film industry. *PloS one* 15 (2): e0229436. doi:10.1371/journal.pone.0229436.

Katz, J. S., and Cothey, V. 2006. Web indicators for complex innovation systems. *Research Evaluation* 15 (2): 85–95. doi:10.3152/147154406781775922.

Kinne, J., and Axenbeck, J. 2018. *Web Mining of Firm Websites : A Framework for Web Scraping and a Pilot Study for Germany.* Mannheim. http://ftp.zew.de/pub/zew-docs/dp/dp18033.pdf%0ADie.

———. 2020. Web mining for innovation ecosystem mapping: A framework and a large-scale pilot study. *Scientometrics* 28 (4): 443. doi:10.1007/s11192-020-03726-9.

Kinne, J., and Lenz, D. 2019. *Predicting Innovative Firms Using Web Mining and Deep Learning.* ZEW Discussion Paper. Mannheim: ZEW.

———. 2021. Predicting innovative firms using web mining and deep learning. *PloS one* 16 (4): e0249071. doi:10.1371/journal.pone.0249071.

Krackhardt, D. 1992. The Strength of Strong Ties: The Importance of Philos in Organizations. In *Networks and Organizations: Structure, Form, and Action*, ed. N. Nohria and R. G. Eccles, 216–39. Boston, MA: Harvard Business School Press.

Krackhardt, D. 1999. The Ties That Torture: Simmelian Tie Analysis in Organizations. *Research in the Sociology of Organizations* 16: 183–210.

Krackhardt, D., and Stern, R. N. 1988. Informal Networks and Organizational Crises: An Experimental Simulation. *Social Psychology Quarterly* 51 (2): 123. doi:10.2307/2786835.

Krüger, M., Kinne, J., Lenz, D., and Resch, B. 2020. *The Digital Layer: How Innovative Firms Relate on the Web.* 20 ZEW Discussion Paper. Mannheim: ZEW.

Lazzeretti, L., and Capone, F. 2016. How proximity matters in innovation networks dynamics along the cluster evolution. A study of the high technology applied to cultural goods. *Journal of Business Research* 69 (12): 5855–65. doi:10.1016/j.jbusres.2016.04.068.

Lingo, E. L., and O'Mahony, S. 2010. Nexus Work: Brokerage on Creative Projects. *Administrative Science Quarterly* 55 (1): 47–81. doi:10.2189/asqu.2010.55.1.47.

Lobo, J., and Strumsky, D. 2008. Metropolitan patenting, inventor agglomeration and social networks: A tale of two effects. *Journal of Urban Economics* 63 (3): 871–84. doi:10.1016/j.jue.2007.07.005.

Manning, C. D., Raghavan P, and Schutze, H. 2009. *An Introduction to Information Retrieval.* Cambridge (Online edition): Cambridge University Press.

Marshall, A. 1920. *Principles of Economics.* London: MacMillan.

Mazzucato, M. 2018. Mission-oriented innovation policies: Challenges and opportunities. *Industrial and Corporate Change* 27 (5): 803–15. doi:10.1093/icc/dty034.

Mewes, L. 2019. Scaling of Atypical Knowledge Combinations in American Metropolitan Areas from 1836 to 2010. *Economic Geography* 95 (4): 341–61. doi:10.1080/00130095.2019.1567261.

Mitchell, M. 2009. *Complexity: A Guided Tour.* Oxford: Oxford University Press.

Molina-Morales, X., Belso-Martínez, J. A., Más-Verdú, F., and Martínez-Cháfer, L. 2015. Formation and dissolution of inter-firm linkages in lengthy and stable networks in clusters. *Journal of Business Research* 68 (7): 1557–62. doi:10.1016/j.jbusres.2015.01.051.

Moody, J. 2011. Network Dynamics. In *The Oxford Handbook of Analytical Sociology*, ed. P. Bearman, P. Hedström, P. Dodds, and D. J. Watts, 447–74: Oxford University Press.

Muscoloni, A., Thomas, J. M., Ciucci, S., Bianconi, G., and Cannistraci, C. V. 2017. Machine learning meets complex networks via coalescent embedding in the hyperbolic space. *Nature communications* 8 (1): 1615. doi:10.1038/s41467-017-01825-5.

Nelson, R. E. 1989. The Strength of Strong Ties: Social Networks and Intergroup Conflict in Organizations. *Academy of Management Journal* 32 (2): 377–401. doi:10.5465/256367.

Nooteboom, B. 2000. Learning by Interaction: Absorptive Capacity, Cognitive Distance and Governance. *Journal of Management and Governance* 4 (1/2): 69–92. doi:10.1023/A:1009941416749.

Obstfeld, D. 2016. Social Networks, the Tertius Iungens Orientation, and Involvement in Innovation. *Administrative Science Quarterly* 50 (1): 100–130. doi:10.2189/asqu.2005.50.1.100.

OECD. 2018. *Oslo Manual: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition, The Measurement of Scientific, Technological and Innovation Activities.* Paris: OECD Publishing.

Oinas, P. 1999. Activity-specificity in organizational learning: implications for analysing the role of proximity. *GeoJournal* 49 (4): 363–72. doi:10.1023/A:1007184012189.

Owen-Smith, J., and Powell, W. W. 2004. Knowledge Networks as Channels and Conduits: The Effects of Spillovers in the Boston Biotechnology Community. *Organization Science* 15 (1): 5–21. doi:10.1287/orsc.1030.0054.

Park, H. W. 2003. Hyperlink network analysis: A new method for the study of social structure on the web. *Connections* 25: 49–61.

Reagans, R., and Zuckerman, E. W. 2001. Networks, Diversity, and Productivity: The Social Capital of Corporate R&D Teams. *Organization Science* 12 (4): 502–17. doi:10.1287/orsc.12.4.502.10637.

Rogers, R. 2010. Mapping Public Web Space with the Issuecrawler. In *Digital Cognitive Technologies: Epistemology and Knowledge Society*, ed. C. Brossard and B. Reber, 115–26. London: Wiley.

Schilling, M. A., and Phelps, C. C. 2007. Interfirm Collaboration Networks: The Impact of Large-Scale Network Structure on Firm Innovation. *Management Science* 53 (7): 1113–26. doi:10.1287/mnsc.1060.0624.

Singh, J. 2005. Collaborative Networks as Determinants of Knowledge Diffusion Patterns. *Management Science* 51 (5): 756–70. doi:10.1287/mnsc.1040.0349.

Storm, H., Baylis, K., and Heckelei, T. 2020. Machine learning in agricultural and applied economics. *European Review of Agricultural Economics* 47 (3): 849–92. doi:10.1093/erae/jbz033.

Strumsky, D., and Lobo, J. 2015. Identifying the sources of technological novelty in the process of invention. *Research Policy* 44 (8): 1445–61. doi:10.1016/j.respol.2015.05.008.

Ter Wal, A. 2014. The dynamics of the inventor network in German biotechnology: Geographic proximity versus triadic closure. *Journal of Economic Geography* 14 (3): 589–620. doi:10.1093/jeg/lbs063.

Ter Wal, A. L. J., Alexy, O., Block, J., and Sandner, P. G. 2016. The Best of Both Worlds: The Benefits of Open-specialized and Closed-diverse Syndication Networks for New Ventures' Success. *Administrative Science Quarterly* 61 (3): 393–432. doi:10.1177/0001839216637849.

Torre, A., and Rallet, A. 2005. Proximity and Localization. *Regional Studies* 39 (1): 47–59. doi:10.1080/0034340052000320842.

Tortoriello, M., and Krackhardt, D. 2010. Activating Cross-Boundary Knowledge: The Role of Simmelian Ties in the Generation of Innovations. *Academy of Management Journal* 53 (1): 167–81. doi:10.5465/amj.2010.48037420.

Tóth, G., and Lengyel, B. 2019. Inter-firm inventor mobility and the role of co-inventor networks in producing high-impact innovation. *The Journal of Technology Transfer* 72 (1): 1. doi:10.1007/s10961-019-09758-5.

Uzzi, B. 1997. Social structure and competition in interfirm networks: The paradox of embeddedness. *Administrative Science Quarterly* 42: 35–67.

Uzzi, B., and Spiro, J. 2005. Collaboration and Creativity: The Small World Problem. *American Journal of Sociology* 111 (2): 447–504.

van der Wouden, F. 2020. A history of collaboration in US invention: Changing patterns of co-invention, complexity and geography. *Industrial and Corporate Change* 29 (3): 599–619. doi:10.1093/icc/dtz058.

van der Wouden, F., and Rigby, D. L. 2019. Co-inventor networks and knowledge production in specialized and diversified cities. *Papers in Regional Science* 98 (4): 1833–53. doi:10.1111/pirs.12432.

van Hoorn, A., and Maseland, R. 2010. Cultural differences between East and West Germany after 1991: Communist values versus economic performance? *Journal of Economic Behavior & Organization* 76 (3): 791–804. doi:10.1016/j.jebo.2010.10.003.

Vaughan, L., Gao, Y., and Kipp, M. 2006. Why are hyperlinks to business Websites created? A content analysis. *Scientometrics* 67 (2): 291–300. doi:10.1007/s11192-006-0100-6.

Vaughan, L., and Wu, G. 2004. Links to commercial websites as a source of business information. *Scientometrics* 60 (3): 487–96. doi:10.1023/B:SCIE.0000034389.14825.bc.

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

Wenau, G., Grigoriev, P., and Shkolnikov, V. 2019. Socioeconomic disparities in life expectancy gains among retired German men, 1997-2016. *Journal of epidemiology and community health* 73 (7): 605–11. doi:10.1136/jech-2018-211742.

West, G. B. 2017. *Scale: The universal laws of life, growth, and death in organisms, cities, and companies.* New York: Penguin Books.

Whittle, A., Lengyel, B., and Kogler, D. F. 2020. Understanding Regional Branching Knowledge Diversification via Inventor Collaboration Networks. *Papers in Evolutionary Economic Geography* 20.06.

Wu, L., Waber, B. N., Aral, S., Brynjolfsson, E., and Pentland, A. 2008. Mining Face-to-Face Interaction Networks using Sociometric Badges: Predicting Productivity in an IT Configuration Task. *SSRN Electronic Journal.* doi:10.2139/ssrn.1130251.

Yang, Z., Algesheimer, R., and Tessone, C. J. 2016. A Comparative Analysis of Community Detection Algorithms on Artificial Networks. *Scientific reports* 6:30750.

Zandbergen, P. A. 2008. A comparison of address point, parcel and street geocoding techniques. *Computers, Environment and Urban Systems* 32 (3): 214–32. doi:10.1016/j.compenvurbsys.2007.11.006.

Appendix A

Statistic	N	Mean	St. Dev.	Min	Pctl(25)	Pctl(75)	Max
INNOVATIVE	432,248	0.109	0.312	0	0	0	1
SIR	432,248	0	1	-2.973	-0.352	-0.352	2.269
SIC	432,248	0	1	-2.824	0.246	0.246	3.317
WIR	432,248	0	1	-1.797	-0.402	0.994	0.994
WIC	432,248	0	1	-1.011	-1.011	0.789	2.589
DEGREE	432,248	1.422	0.704	0.693	0.693	1.792	11.852
ALTER	432,248	6.152	5.186	0	0	11.066	18.955
TRANSITIVITY	432,248	0.107	0.226	0	0	0.1	1
AGE	405,514	2.891	0.859	0.651	2.307	3.400	6.928
METROPOL	432,248	0.718	0.450	0	0	1	1
EAST	432,248	0.181	0.385	0	0	0	1
FDENSITY	432,248	3.951	1.585	0	2.944	4.934	8.277
RFDENSITY	432,248	9.890	0.574	8.411	9.448	10.515	10.705

Note: Isolated firms (201,275) were removed from the analysis. A small share of firms (26,734) does not include age information.

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)
(1) INNOVATIVE	1												
(2) SIR	0.09	1											
(3) SIC	0.03	-0.06	1										
(4) WIR	0.04	-0.03	0.1	1									
(5) WIC	0.03	0.17	-0.16	-0.14	1								
(6) DEGREE	0.1	0.27	-0.14	-0.2	0.38	1							
(7) ALTER	0.03	-0.14	0.21	0.48	-0.1	0.07	1						
(8) TRANSITIVITY	0	0.3	-0.45	-0.2	0.31	0.08	-0.44	1					
(9) AGE	-0.1	-0.01	-0.03	-0.07	0.04	0.12	-0.02	-0.01	1				
(10) METROPOL	0.07	0	0	-0.02	0.01	0.03	0	-0.01	-0.07	1			
(11) EAST	-0.02	-0.02	-0.02	-0.01	0	0.01	-0.03	0.01	-0.05	-0.01	1		
(12) FDENSITY	0.12	0.01	-0.02	-0.05	0.04	0.11	-0.02	0.01	-0.07	0.35	0	1	
(13) RFDENSITY	0.08	-0.05	-0.01	-0.07	0	0.02	0.01	-0.01	-0.04	0.27	-0.18	0.28	1

Appendix B

	Dependent variable: INNOVATIVE (binarized based on the median value of InnoProb)					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR		0.2732***				0.2696***
		(0.0128)				(0.0130)
WIR			0.3155***			0.3040***
			(0.0090)			(0.0091)
SIC				0.0211*		-0.0025
				(0.0121)		(0.0122)
WIC					-0.0868***	-0.0475***
					(0.0078)	(0.0080)
DEGREE	0.2594***	0.2432***	0.3435***	0.2565***	0.2746***	0.3332***
	(0.0053)	(0.0053)	(0.0058)	(0.0055)	(0.0055)	(0.0062)
ALTER	0.0002	0.0010	-0.0131***	0.0003	-0.0006	-0.0122***
	(0.0007)	(0.0007)	(0.0008)	(0.0007)	(0.0007)	(0.0008)
TRANSITIVITY	-0.0971***	-0.2311***	-0.0144	-0.0984***	-0.0300*	-0.1126***
	(0.0169)	(0.0181)	(0.0171)	(0.0169)	(0.0179)	(0.0191)
AGE	-0.3063***	-0.3038***	-0.3018***	-0.3062***	-0.3066***	-0.2996***
	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)	(0.0042)
FDENSITY	0.1117***	0.1134***	0.1133***	0.1117***	0.1116***	0.1148***
	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)	(0.0025)
RFDENSITY	0.0493***	0.0586***	0.0694***	0.0495***	0.0491***	0.0777***
	(0.0064)	(0.0064)	(0.0065)	(0.0064)	(0.0064)	(0.0065)
METROPOL	0.0466***	0.0457***	0.0444***	0.0465***	0.0460***	0.0433***
	(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)	(0.0083)
EAST	-0.1686***	-0.1648***	-0.1665***	-0.1684***	-0.1691***	-0.1631***
	(0.0091)	(0.0091)	(0.0091)	(0.0091)	(0.0091)	(0.0091)
Constant	-0.8267***	-0.9176***	-1.2034***	-0.8266***	-0.8084***	-1.2695***
	(0.0718)	(0.0719)	(0.0727)	(0.0718)	(0.0718)	(0.0729)
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-247,131.5000	-246,903.8000	-246,509.3000	-247,130.0000	-247,070.2000	-246,274.9000
Akaike Inf. Crit.	494,363.0000	493,909.7000	493,120.6000	494,362.0000	494,242.5000	492,657.7000
Note:						*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: INNOVATIVE (binarized based on the median value of InnoProb)					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR (dummy)	0.3047***	0.2388***	0.3096***			
	(0.0150)	(0.0141)	(0.0249)			
SIC (dummy)	0.0219			-0.0328**	0.0396*	
	(0.0140)			(0.0132)	(0.0204)	
WIR (dummy)		0.3050***		0.2968***		0.3043***
		(0.0091)		(0.0091)		(0.0105)
WIC (dummy)			-0.0882***		-0.0846***	-0.0467***
			(0.0081)		(0.0083)	(0.0102)
SIR×SIC	-0.1183***					
	(0.0289)					
SIR×WIR		0.1556***				
		(0.0322)				
SIR×WIC			-0.0436			
			(0.0285)			
SIC×WIR				0.3446***		
				(0.0318)		
SIC×WIC					-0.0237	
					(0.0248)	
WIR×WIC						0.0100
						(0.0161)
Constant	-0.9161***	-1.2741***	-0.8981***	-1.1709***	-0.8073***	-1.1827***
	(0.0719)	(0.0729)	(0.0720)	(0.0727)	(0.0718)	(0.0729)
Controls	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-246,895.3000	-246,280.8000	-246,835.3000	-246,447.6000	-247,067.8000	-246,494.7000
Akaike Inf. Crit.	493,896.6000	492,667.7000	493,776.6000	493,001.1000	494,241.7000	493,095.4000
Note:						*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: INNOVATIVE (binarized based on the 75th percentile of InnoProb)					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR		0.1473***				0.1268***
		(0.0044)				(0.0045)
WIR			0.1864***			0.1750***
			(0.0049)			(0.0049)
SIC				0.0641***		0.0596***
				(0.0043)		(0.0044)
WIC					-0.0260***	-0.0213***
					(0.0044)	(0.0045)
DEGREE	0.3318***	0.2675***	0.3876***	0.3405***	0.3441***	0.3465***
	(0.0057)	(0.0060)	(0.0059)	(0.0057)	(0.0060)	(0.0067)
ALTER	0.0012	0.0029***	-0.0157***	0.0009	0.0011	-0.0136***
	(0.0008)	(0.0008)	(0.0010)	(0.0008)	(0.0008)	(0.0010)
TRANSITIVITY	-0.0450**	-0.2353***	-0.0473**	0.0721***	-0.0137	-0.0754***
	(0.0198)	(0.0209)	(0.0198)	(0.0213)	(0.0205)	(0.0232)
AGE	-0.3372***	-0.3319***	-0.3315***	-0.3357***	-0.3373***	-0.3258***
	(0.0048)	(0.0049)	(0.0049)	(0.0048)	(0.0048)	(0.0049)
FDENSITY	0.1318***	0.1350***	0.1338***	0.1319***	0.1318***	0.1366***
	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)	(0.0029)
RFDENSITY	0.1100***	0.1248***	0.1323***	0.1113***	0.1098***	0.1447***
	(0.0074)	(0.0075)	(0.0075)	(0.0074)	(0.0074)	(0.0075)
METROPOL	0.1013***	0.0977***	0.0967***	0.1012***	0.1010***	0.0936***
	(0.0099)	(0.0100)	(0.0100)	(0.0099)	(0.0099)	(0.0100)
EAST	-0.2005***	-0.1884***	-0.1974***	-0.1982***	-0.2008***	-0.1854***
	(0.0108)	(0.0109)	(0.0109)	(0.0108)	(0.0108)	(0.0109)
Constant	-3.2208***	-3.2764***	-3.4202***	-3.2620***	-3.2394***	-3.5082***
	(0.0930)	(0.0932)	(0.0933)	(0.0931)	(0.0931)	(0.0936)
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-197,579.6000	-197,009.4000	-196,836.4000	-197,469.6000	-197,562.3000	-196,277.1000
Akaike Inf. Crit.	395,259.2000	394,120.8000	393,774.8000	395,041.2000	395,226.6000	392,662.1000
Note:						*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: INNOVATIVE (binarized based on the 75th percentile of InnoProb)					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR (dummy)	0.2851***	0.2395***	0.3665***			
	(0.0162)	(0.0156)	(0.0251)			
SIC (dummy)	0.0853***			0.0416***	0.1242***	
	(0.0154)			(0.0147)	(0.0213)	
WIR (dummy)		0.3276***		0.3182***		0.3395***
		(0.0105)		(0.0105)		(0.0121)
WIC (dummy)			-0.0532***		-0.0557***	-0.0107
			(0.0092)		(0.0095)	(0.0119)
SIR×SIC	-0.0291					
	(0.0307)					
SIR×WIR		0.1666***				
		(0.0317)				
SIR×WIC			-0.1086***			
			(0.0294)			
SIC×WIR				0.3584***		
				(0.0315)		
SIC×WIC					-0.0244	
					(0.0265)	
WIR×WIC						0.0054
						(0.0184)
Constant	-3.3244***	-3.6961***	-3.3070***	-3.5775***	-3.2046***	-3.6126***
	(0.0932)	(0.0942)	(0.0933)	(0.0940)	(0.0931)	(0.0942)
Controls	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-197,350.6000	-196,805.7000	-197,337.5000	-196,915.9000	-197,525.0000	-197,016.5000
Akaike Inf. Crit.	394,807.2000	393,717.3000	394,781.0000	393,937.7000	395,156.0000	394,138.9000
Note:						*p<0.1; **p<0.05; ***p<0.01

Appendix C

	Dependent variable: INNOVATIVE (based on the patent status)					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR		0.0792^{***}				0.0708^{***}
		(0.0135)				(0.0137)
WIR			0.0675^{***}			0.0714^{***}
			(0.0161)			(0.0164)
SIC				0.0873^{***}		0.0819^{***}
				(0.0128)		(0.0129)
WIC					0.0680^{***}	0.0712^{***}
					(0.0134)	(0.0135)
DEGREE	0.6511^{***}	0.6126^{***}	0.6591^{***}	0.6642^{***}	0.6354^{***}	0.6204^{***}
	(0.0148)	(0.0163)	(0.0149)	(0.0150)	(0.0152)	(0.0172)
ALTER	0.0046^{*}	0.0045^{*}	-0.0009	0.0042	0.0051^{*}	-0.0009
	(0.0027)	(0.0027)	(0.0030)	(0.0027)	(0.0027)	(0.0030)
TRANSITIVITY	0.1354^{**}	0.0251	0.1469^{**}	0.2840^{***}	0.0585	0.1080
	(0.0666)	(0.0702)	(0.0666)	(0.0699)	(0.0687)	(0.0763)
AGE	0.4426^{***}	0.4453^{***}	0.4456^{***}	0.4439^{***}	0.4421^{***}	0.4489^{***}
	(0.0165)	(0.0165)	(0.0165)	(0.0165)	(0.0165)	(0.0165)
FDENSITY	-0.0314^{***}	-0.0294^{***}	-0.0304^{***}	-0.0318^{***}	-0.0321^{***}	-0.0296^{***}
	(0.0101)	(0.0101)	(0.0101)	(0.0101)	(0.0101)	(0.0101)
RFDENSITY	-0.0290	-0.0201	-0.0209	-0.0270	-0.0278	-0.0096
	(0.0247)	(0.0247)	(0.0248)	(0.0247)	(0.0247)	(0.0249)
METROPOL	-0.0569^{*}	-0.0608^{**}	-0.0593^{**}	-0.0568^{*}	-0.0566^{*}	-0.0622^{**}
	(0.0301)	(0.0301)	(0.0301)	(0.0301)	(0.0301)	(0.0301)
EAST	0.0522	0.0599^{*}	0.0558	0.0528	0.0503	0.0610^{*}
	(0.0349)	(0.0349)	(0.0349)	(0.0349)	(0.0349)	(0.0349)
Constant	-8.1773^{***}	-8.1960^{***}	-8.2364^{***}	-8.2325^{***}	-8.1615^{***}	-8.2922^{***}
	(0.4004)	(0.4005)	(0.4008)	(0.4006)	(0.4005)	(0.4012)
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-26,298.8100	-26,281.5000	-26,289.8800	-26,275.6800	-26,286.0700	-26,239.1500
Akaike Inf. Crit.	52,697.6300	52,665.0000	52,681.7500	52,653.3500	52,674.1500	52,586.3000
Note:						*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: INNOVATIVE (based on the patent status)					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR (dummy)	-0.0096	-0.0623	0.2551***			
	(0.0506)	(0.0500)	(0.0624)			
SIC (dummy)	0.1707***			0.0427	0.3407***	
	(0.0420)			(0.0431)	(0.0542)	
WIR (dummy)		-0.0229		-0.0556		-0.0194
		(0.0336)		(0.0341)		(0.0394)
WIC (dummy)			0.1206***		0.1188***	0.0320
			(0.0294)		(0.0306)	(0.0374)
SIR×SIC	0.1333					
	(0.0867)					
SIR×WIR		0.3765***				
		(0.0831)				
SIR×WIC			-0.3245***			
			(0.0805)			
SIC×WIR				0.5794***		
				(0.0757)		
SIC×WIC					-0.2434***	
					(0.0714)	
WIR×WIC						0.1489**
						(0.0586)
Constant	-8.1992***	-8.1668***	-8.2445***	-8.1068***	-8.2145***	-8.2219***
	(0.4009)	(0.4026)	(0.4012)	(0.4023)	(0.4008)	(0.4029)
Controls	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-26,281.4600	-26,287.5500	-26,284.9400	-26,254.3200	-26,273.3600	-26,290.1600
Akaike Inf. Crit.	52,668.9100	52,681.1000	52,675.8700	52,614.6500	52,652.7200	52,686.3300
Note:						*p<0.1; **p<0.05; ***p<0.01

Appendix D

	Dependent variable: INNOVATIVE					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR		0.1839***				0.1606***
		(0.0061)				(0.0062)
WIR			0.2045***			0.1901***
			(0.0069)			(0.0070)
SIC				0.0876***		0.0813***
				(0.0058)		(0.0059)
WIC					-0.0031	0.0039
					(0.0060)	(0.0061)
DEGREE	0.3605***	0.2742***	0.4124***	0.3702***	0.3618***	0.3395***
	(0.0073)	(0.0079)	(0.0075)	(0.0074)	(0.0077)	(0.0086)
ALTER	0.0057***	0.0077***	-0.0115***	0.0053***	0.0057***	-0.0090***
	(0.0012)	(0.0012)	(0.0013)	(0.0012)	(0.0012)	(0.0013)
TRANSITIVITY	0.0739***	-0.1871***	0.0965***	0.2280***	0.0776***	0.0061
	(0.0277)	(0.0298)	(0.0278)	(0.0295)	(0.0286)	(0.0326)
AGE	-0.3621***	-0.3567***	-0.3566***	-0.3600***	-0.3621***	-0.3500***
	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0068)	(0.0069)
FDENSITY	0.1586***	0.1636***	0.1611***	0.1586***	0.1586***	0.1652***
	(0.0040)	(0.0041)	(0.0041)	(0.0040)	(0.0040)	(0.0041)
RFDENSITY	0.1888***	0.2068***	0.2110***	0.1909***	0.1888***	0.2269***
	(0.0105)	(0.0106)	(0.0106)	(0.0105)	(0.0105)	(0.0106)
METROPOL	0.1913***	0.1846***	0.1844***	0.1908***	0.1913***	0.1789***
	(0.0148)	(0.0148)	(0.0148)	(0.0148)	(0.0148)	(0.0148)
EAST	-0.1695***	-0.1550***	-0.1675***	-0.1659***	-0.1695***	-0.1523***
	(0.0153)	(0.0154)	(0.0154)	(0.0154)	(0.0153)	(0.0154)
Constant	-5.9594***	-6.0129***	-6.1531***	-6.0160***	-5.9612***	-6.2329***
	(0.1676)	(0.1678)	(0.1679)	(0.1677)	(0.1676)	(0.1682)
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-114,082.6000	-113,609.9000	-113,630.4000	-113,970.5000	-114,082.5000	-113,150.9000
Akaike Inf. Crit.	228,265.2000	227,321.7000	227,362.8000	228,043.0000	228,266.9000	226,409.8000
Note:						*p<0.1; **p<0.05; ***p<0.01

	Dependent variable: INNOVATIVE					
	(1)	(2)	(3)	(4)	(5)	(6)
SIR (dummy)	0.3130***	0.2548***	0.4127***			
	(0.0217)	(0.0212)	(0.0311)			
SIC (dummy)	0.1574***			0.0882***	0.1702***	
	(0.0200)			(0.0195)	(0.0269)	
WIR (dummy)		0.3251***		0.3105***		0.3518***
		(0.0145)		(0.0146)		(0.0169)
WIC (dummy)			-0.0134		-0.0263**	0.0270
			(0.0128)		(0.0132)	(0.0166)
SIR×SIC	-0.0372					
	(0.0395)					
SIR×WIR		0.1955***				
		(0.0390)				
SIR×WIC			-0.1341***			
			(0.0373)			
SIC×WIR				0.4159***		
				(0.0378)		
SIC×WIC					0.0168	
					(0.0339)	
WIR×WIC						0.0040
						(0.0253)
Constant	-6.0857***	-6.4398***	-6.0784***	-6.2920***	-5.9547***	-6.3630***
	(0.1679)	(0.1688)	(0.1679)	(0.1686)	(0.1677)	(0.1688)
Controls	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-113,894.5000	-113,624.2000	-113,923.1000	-113,656.4000	-114,025.8000	-113,773.2000
Akaike Inf. Crit.	227,895.0000	227,354.5000	227,952.2000	227,418.7000	228,157.5000	227,652.4000
Note:						*p<0.1; **p<0.05; ***p<0.01

Appendix E

	Dependent variable: INNOVATIVE										
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
SIR	0.3130***	0.2548***	0.4127***				0.2486***	0.3768***	0.3567***		0.3381***
	(0.0217)	(0.0212)	(0.0311)				(0.0248)	(0.0387)	(0.0496)		(0.0627)
SIC	0.1574***			0.0882***	0.1702***		0.0683***	0.1132***		-0.0035	-0.0417
	(0.0200)			(0.0195)	(0.0269)		(0.0226)	(0.0309)		(0.0358)	(0.0392)
WIR		0.3251***		0.3105***		0.3518***	0.2953***		0.3319***	0.3022***	0.2899***
		(0.0145)		(0.0146)		(0.0169)	(0.0150)		(0.0174)	(0.0176)	(0.0180)
WIC			-0.0134		-0.0263**	0.0270		-0.0260*	0.0271	0.0055	0.0034
			(0.0128)		(0.0132)	(0.0166)		(0.0136)	(0.0172)	(0.0181)	(0.0186)
SIR×SIC	-0.0372						0.0114	0.0422			0.1170
	(0.0395)						(0.0460)	(0.0678)			(0.1054)
SIR×WIR		0.1955***					0.1747***		0.0368		0.0173
		(0.0390)					(0.0469)		(0.0624)		(0.0781)
SIR×WIC			-0.1341***					-0.0883*	-0.1256**		-0.1010
			(0.0373)					(0.0453)	(0.0543)		(0.0676)
SIC×WIR				0.4159***			0.4285***			0.4737***	0.5104***
				(0.0378)			(0.0447)			(0.0517)	(0.0604)
SIC×WIC					0.0168			0.0741*		0.1287***	0.1626***
					(0.0339)			(0.0390)		(0.0426)	(0.0476)
WIR×WIC						0.0040			-0.0024	0.0234	0.0186
						(0.0253)			(0.0263)	(0.0268)	(0.0276)
SIR×SIC×WIR							-0.2530***				-0.3227**
							(0.0884)				(0.1381)
SIR×SIC×WIC								-0.1357			-0.1663
								(0.0825)			(0.1170)

SIR×WIR×WIC									0.3498***		0.3399***
									(0.0902)		(0.1105)
SIC×WIR×WIC										-0.0276	-0.0842
										(0.0804)	(0.0937)
SIR×SIC×WIR×WIC											0.0823
											(0.2047)
Constant	-6.0857***	-6.4398***	-6.0784***	-6.2920***	-5.9547***	-6.3630***	-6.3826***	-6.0739***	-6.4516***	-6.2985***	-6.3857***
	(0.1679)	(0.1688)	(0.1679)	(0.1686)	(0.1677)	(0.1688)	(0.1690)	(0.1679)	(0.1691)	(0.1689)	(0.1692)
Controls	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Sector FE	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	404,857	404,857	404,857	404,857	404,857	404,857	404,857	404,857	404,857	404,857	404,857
Log Likelihood	-113,894.5000	-113,624.2000	-113,923.1000	-113,656.4000	-114,025.8000	-113,773.2000	-113,536.1000	-113,884.4000	-113,614.0000	-113,647.3000	-113,518.6000
Akaike Inf. Crit.	227,895.0000	227,354.5000	227,952.2000	227,418.7000	228,157.5000	227,652.4000	227,186.2000	227,882.9000	227,342.0000	227,408.7000	227,167.3000
Note:											*p<0.1; **p<0.05; ***p<0.01

//

IMPRINT

ZEW – Leibniz-Zentrum für Europäische Wirtschaftsforschung GmbH Mannheim

ZEW – Leibniz Centre for European Economic Research

L 7,1 · 68161 Mannheim · Germany
Phone +49 621 1235-01
info@zew.de · zew.de