Firm Matching in the Market for Technology

Pere Arqué-Castells* and Daniel F. Spulber†
Northwestern University

April 20, 2017

Abstract

To portray the market for technology, we create a dataset that tracks interactions in the market for technology between publicly held companies in North America with at least one patent. The dataset offers a broad coverage over time (1990-2013) and across sectors, technologies and contractual forms of exchange (i.e. patent trades, licensing, cross-licensing and R&D alliances). Using the dataset, we study firm matching in the market for technology as a function of three metrics that have been widely documented as important catalysts of technology adoption and knowledge spillovers: market, technological and geographical proximity. We predict that proximity (in any of these three dimensions) will have a positive effect on the probability of a match in the market (but not necessarily on adoption through infringement) based on a model of technology transfer between a provider and an adopter. The three proximity metrics are found to have a positive and significant effect on the probability of a match when exploiting between adopter-provider variation. Only technological proximity remains positive and significant when adopter-provider variation is exploited. The results have implications that may be useful in characterizing results in the spillovers literature.

Keywords: Market for technology; Patents; Spillovers; Innovation

JEL Codes: O31; O34

*Email: pere.arque@law.northwestern.edu. Research Associate at the Searle Center on Law, Regulation and Economic Growth. Support from Qualcomm is gratefully acknowledged.

†Elinor Hobbs Distinguished Professor of International Business, Professor of Strategy, Kellogg School of Management, Northwestern University, 2211 University Drive, Evanston, IL, 60208. E-mail: jems@kellogg.northwestern.edu. I gratefully acknowledge research support from the Ewing Marion Kauffman Foundation, Qualcomm, and the Kellogg School of Management. The views expressed are those of the authors and do not necessarily represent those of any of these institutions.

1 Introduction

Most exchanges of abstract knowledge take place through bilateral bargaining (Gans & Stern, 2010, Arora & Gambardella, 2010, Spulber, 2016). This implies that the market for technology is essentially the web of bilateral interactions between the actors in the market. In this paper, we portray the market for technology by creating a dataset that tracks interactions between publicly held companies in North America with at least one patent in the USPTO Patent Assignment Dataset. This dataset includes 22,247 unique interactions between 4,707 firms participating in the market for technology. While not a census, it compiles the most comprehensive sample of interactions in the market for technology that can be put together for public firms. Moreover, it offers a broad coverage over time (1990-2013) and across sectors, technologies and contractual forms of exchange (i.e. patent trades, licensing, cross-licensing and R&D alliances).

The constructed dataset is naturally well suited to study firm matching in the market for technology. The market for technology has the potential to generate social gains by enhancing allocative efficiency and promoting the specialization of innovative labor (Arora & Gambardella, 2010). However, this potential can only be materialized if the market for technology is efficient at matching each idea with the firm best fit to commercialize it (Gans & Stern, 2010; Akcigit *et al.* , 2016). Studying matching is important to understand where the market works and where it does not.

We study matching as a function of three distance metrics that have been widely documented as important catalysts of technology adoption and knowledge spillovers: market, technological and geographical proximity (see Lychagin *et al.* , 2016; Hall *et al.* , 2010). If the profits from knowledge adoption are increasing in such metrics so should the gains from trade and the probability of a match in the market for technology. Understanding the precise instances in which adoption takes place on and off the market is important to separate pure externalities from transfers internalized by technology providers.

To analyze the impact of the proximity metrics (i.e. market, technological and geographical proximity) on the probability of a match, a model of knowledge transfer between a technology provider and a technology adopter is built. The technology provider owns a patented technology that can be transferred to the adopter through negotiation of a patent licensing agreement. Alternatively, the adopter can adopt the technology by infringing on the provider's patent. The profits from technology adoption are increasing in (market, technological or geographical) proximity. The main result is that the probability of a match is unambiguously increasing in proximity because the gains from trade are also increasing in proximity. On the other hand, the probability of adoption through infringement (arguably the form of adoption conductive to pure knowledge spillovers) is not necessarily increasing in proximity.

The dataset on interactions is first analyzed through a descriptive network analysis that provides several interesting stylized facts. First, firms very strongly cluster by sector of activity, technological field and geographical location. Second, the Pharma and the ICT clusters dominate the market. Third, horizontal (between rivals) and vertical (between non-rivals) structures of interaction coexist. Fourth, prominent adopters also tend to be prominent providers of technology.

Next, a proper econometric analysis is carried out to tease out the individual effect of market, technological and geographical proximity on the probability of a match. To this end, the datastet on interactions is complemented with Compustat and USPTO Patent Assigment data to define the distance metrics and a set of controls. Market, technological and geographical proximity between firms are constructed as the Jaffe (1986) uncentered correlation closeness measure using Compustat Segment sales, USPTO classes and USPTO inventor counties respectively (similarly to Bloom *et al.* , 2013). Three different datasets are used in the regressions. Dataset A, which collapses all the years to create a cross section of adopter-provider pairings; Dataset B, which splits the sample into two periods to create predetermined explanatory variables; and Dataset C, which has a longitudinal structure that allows controlling for adopter-provider fixed effects.

The main empirical results are as follows. First, all the proximity metrics are found to positively affect the probability of a match when exploiting the between adopter-provider variation in Datasets A to C. This is true for every contractual form of exchange ranging from patent trades to R&D alliances. Within adopter-provider estimates on Dataset C are less unanimous with technological and geographical proximity having a positive (negative) effect on the probability of a match for non-alliance (within-alliance) interactions, and market proximity being no longer significant. On top of that, the probability of a match is found to be monotonically increasing in technological and geographical proximity for (almost) every single form of exchange. A different pattern emerges for market proximity with the probability of a match immediately raising above zero for any positive value, but remaining fairly constant as proximity increases. Overall, the results for market proximity suggest a tension between the rent creation and rent dissipation effects of market proximity (see Arora & Fosfuri, 2003).

Related literature. Our paper relates to the vast literature on the market for technology (Arora *et al.* , 2001; Arora & Gambardella, 2010, Spulber, 2013; Spulber, 2015; Spulber, 2016). An important branch of the literature focuses on the supply side of the market, placing a great emphasis on the licensor's dilemma between maximizing licensing revenues and minimizing rent dissipation from increased competition (see Arora & Fosfuri, 2003; Gambardella & Giarratana, 2013; Fosfuri, 2006; Gambardella *et al.* , 2007).[1] We contribute to this literature by pointing out that, besides triggering rent dissipation on the provider, market proximity may also have a positive rent creation effect on the profits of the adopter.[2] The overall effect of market proximity on the probability of a match is then the result of these two countervailing forces. Our estimates indicate that market proximity either has a

[1] Arora & Fosfuri (2003) conclude that licensing to competitors is more likely if the downstream market is more competitive because in such a case the licensee dissipates fewer rents. Fosfuri (2006) concludes that licensing to competitors is also more likely if the market for technology is more competitive because if the licensor does not grant the license himself someone else will.

[2] This assumption is in keeping with the spillovers literature (see Bernstein & Nadiri, 1988). Such rent creation effect may be due to technologies being tailored to specific market needs or to the adopter's better understanding of the market.

positive effect on the probability of a match or no effect at all, suggesting that rent creation may dominate rent dissipation.

Some recent papers study the demand side of the market for technology (see Ceccagnoli *et al.* , 2010; Ali & Cockburn, 2016). Such papers emphasize that idea complementarity is a key determinant of demand for two reasons. First, it helps the buyer evaluate the technology and eliminate uncertainty about its value (Cohen & Levinthal, 1989; Cassiman & Veugelers, 2002; Cassiman & Veugelers, 2006). Second, many ideas only have value when matched to key complementary assets (Teece, 1986). We show that the probability of a match is indeed increasing in technological proximity (a proxy for complementarity). This is in line with Ali & Cockburn (2016) who find licensees to prefer patents that are technologically closer to their portfolio.

Like us, a few recent papers study the two sides of the market. Figueroa & Serrano (2013) study the patterns of patent trading flows of small and large firms, finding the patent fit with the firms' respective portfolios to be an important determinant of patent sale and acquisition decisions.[3] Akcigit *et al.* (2016) document similar stylized facts for publicly traded firms in North America. Drivas & Economidou (2015) and Drivas *et al.* (2016) find geographical proximity to matter for patent trades. An important difference between these papers and ours is that their matching is patent-to-firm or patent-to-patent while ours is firm-to-firm. This allows us to actually study firm matching in the market for technology.

Our paper very closely relates to the literature on R&D spillovers. It is well known that R&D externalities are prominent between firms in similar industries (Teece, 1986; Bernstein & Nadiri, 1988), technological fields (Jaffe, 1986; Bloom *et al.* , 2013; Manresa, 2016) and geographical areas (Bottazzi & Peri, 2003). Lychagin *et al.* (2016) jointly study the three types of R&D spillovers. In this paper we show that interactions in the market for technology are not orthogonal to industrial, technological and geographical proximity either. In other words, higher externalities come with more internalization. That should be acknowledged

[3] Sold patents have a low fit with the portfolio of the buyer and a high fit with the portfolio of the seller.

when calculating the wedge between the social and private rates of return to R&D. Arqué-Castells & Spulber (2017) study R&D spillovers accounting for the existence of technology markets.

The remaining of the paper is organized as follows. Section 2 presents the analytical model. Section 3 presents the data. Section 4 presents a visual network analysis of the market for technology. Section 5 discusses the econometric analysis. Section 6 presents the econometric results. Section 7 provides a discussion on the implications of the results for the regression based spillovers literature. Section 8 concludes.

2 The Basic Model

To examine the effects of technological, geographical and market proximity on the probability of a match in the market for technology, it is sufficient to consider the interaction between a technology provider and a technology adopter. The technology provider owns a patented technology that can be transferred to the adopter through negotiation of a patent licensing agreement.[4] Alternatively, the adopter can adopt the technology by infringing on the provider's patent. This setting is sufficient to represent interactions between multiple technology providers and technology adopters.

Let $t = t(x, \theta)$ represent the incremental profit the adopter obtains from the technology transfer where x is a shock that is specific to the provider-adopter pair, and θ represents proximity between the provider and the adopter respectively. *Proximity* θ refers to either technological, geographical, or market distance measures. Assume that the incremental profit from technology transfer is continuously differentiable and increasing in the shock x and proximity θ.[5]

[4] The model takes licensing as the contractual mechanism for technology transfer, but the analysis applies to other forms of market exchange such as patent trades, cross-licensing and R&D alliances.

[5] In defining $t(x, \theta)$ we rely on evidence from the R&D spillovers literature which has consistently documented that technological, geographical and market proximity are important determinants of a firm's abiality to benefit from external knowledge (see Lychagin *et al.* , 2016).

Technology adoption may trigger some rent dissipation on the provider. Let $\lambda \in [0,1)$ be the intensity of rent dissipation and $\lambda t(x,\theta)$ the total reduction in profit to the provider.[6]

The shock x is observed before the provider and the adopter make any technology transfer decisions. Assume that the shock is distributed on the positive real line with cumulative distribution $F(x)$. Technology shocks are independent across provider-adopter pairs. The probability of a successful match between a provider and an adopter will depend on the realization of the shock. This will generate predictions about the effects of technological proximity and market competiton on the probability of a match.

If there is no technology adoption, the profits of the provider and the adopter are given by π_0^P and π_0^A respectively. If adoption takes place through a technology transfer, the provider and the adopter incur licensing costs c^P and c^A respectively and their profits are given by

$$\Pi^P(x,\theta) = \pi_0^P - \lambda t(x,\theta) - c^P, \tag{1}$$

$$\Pi^A(x,\theta) = \pi_0^A + t(x,\theta) - c^A. \tag{2}$$

Consider now the effects of patent infringement by the adopter. If the adopter infringes on the provider's patent suppose that, with probability p, the adopter must cease infringement and pay damages K to the provider. The probability p denotes the likelihood that infringement will be detected and the patent will be found to be valid. Otherwise, the adopter benefits from the infringement and the provider suffers harm with probability $1-p$. If the adopter infringes on the provider's technology, the expected profits of the provider and the adopter are then

$$\Pi_0^P(x,\theta) = (1-p)[\pi_0^P - \lambda t(x,\theta)] + p[\pi_0^P + K], \tag{3}$$

[6]Technological and geographical proximity are likely to enter $t(x,\theta)$ only. On the other hand, market proximity could enter both $t(x,\theta)$ and $\lambda(\theta)$ because higher market proximity may also trigger more rent dissipation on the provider (see Arora & Fosfuri, 2003). The overall effect of market proximity would then depend on the strength of the rent creation and rent dissipation effects. Here we just model the rent creation effect but we shall keep in mind the rent dissipation effect in the discussion of the empirical results for market proximity.

$$\Pi_0^A(x,\theta) = (1-p)[\pi_0^A + t(x,\theta)] + p[\pi_0^A - K]. \tag{4}$$

The outcome of negotiation between the provider and the adopter depends on whether the adopter's outside option is to avoid the technology entirely or to infringe on the technology. The adopter's outside option is infringement if and only if the expected profits from infringement are greater than or equal to profits without technology adoption, $\Pi_0^A(x,\theta) \geq \pi_0^A$. The critical value of the shock for which the adopter is indifferent between infringement and non-adoption is given by a unique value $\widetilde{x} = \widetilde{x}(\theta)$ that solves $\Pi_0^A(\widetilde{x},\theta) = \pi_0^A$, so that

$$t(\widetilde{x},\theta) = \frac{p}{1-p}K. \tag{5}$$

Because $t(x,\theta)$ is increasing in x, if $x \geq \widetilde{x}(\theta)$, the adopter prefers infringement to non-adoption and conversely for $x < \widetilde{x}(\theta)$.

The outside options for the provider and the adopter are defined as follows,

$$\pi^j(x,\theta) = \begin{cases} \Pi_0^j(x,\theta) & \text{if } x \geq \widetilde{x}(\theta), \\ \pi_0^j & \text{if } x < \widetilde{x}(\theta), \end{cases} \qquad j = P, A.$$

This implies that the overall gains from technology licensing equal

$$G(x,\theta) = \Pi_0^P(x,\theta) + \Pi_0^A(x,\theta) - \pi^P(x,\theta) - \pi^A(x,\theta). \tag{6}$$

Patent licensing is feasible if and only if there are gains from technology licensing, $G(x,\theta) \geq 0$. Define $x^* = x^*(\theta)$ as the smallest critical value of the shock such that there are gains from technology licensing, that is, $G(x^*,\theta) = 0$.

Full information bargaining will generate a Pareto efficient outcome. This holds whether bargaining is cooperative or non-cooperative.[7] So, if there are gains from technology licensing, the provider and the adopter negotiate royalties R to divide total surplus. With bargaining over patent licensing royalties, we can characterize the effects of the technology

[7] For axiomatic bargaining see for example Nash (1953). For non-cooperative bargaining see for example Rubinstein (1982).

shock on the outcome.

Proposition 1. *If $x^* \leq \widetilde{x}$, there are two regions, with licensing occurring if $x^* \leq x < \infty$ and no technology adoption if $0 \leq x < x^*$. If $x^* > \widetilde{x}$, there are three regions, with licensing occurring if $x^* \leq x < \infty$, infringement occurring if $\widetilde{x} \leq x < x^*$, and no technology adoption if $0 \leq x < \widetilde{x}$.*

The intuition for this result is as follows. If $x^* \leq \widetilde{x}$, the outside options for the two parties are those with no technology transfer for $x^* \leq x \leq \widetilde{x}$ and the outside options for the two parties are those with infringement for $\widetilde{x} \leq x$. In either situation, licensing generates gains from trade and the parties negotiate a licensing agreement for $x \geq x^*$. If $x^* > \widetilde{x}$, the outside options for the two parties are those with infringement for all $x \geq x^*$ and the parties negotiate a licensing agreement for $x \geq x^*$.

A match occurs when the parties negotiate a licensing agreement so that the probability of a match is given by

$$\Pr(\text{match}) = 1 - F(x^*(\theta)). \tag{7}$$

We are interested in understanding how this probability varies with respect to θ. The proofs of the next two results are in the Appendix.

Proposition 2. *The probability of a match is increasing in proximity θ.*

Proposition 2 is empirically testable. In the next sections we describe the dataset and empirical strategy used to test if the probability of a match in the market for technology is increasing in proximity.

Proposition 1 implies that there is adoption if $x \geq \min\{\widetilde{x}(\theta), x^*(\theta)\}$. Adoption takes place through infringement if $\widetilde{x} \leq x < x^*$ or through licensing if $x^* \leq x$. This implies that the probability of adoption is given by

$$\Pr(\text{adoption}) = 1 - F(\min\{\widetilde{x}(\theta), x^*(\theta)\}). \tag{8}$$

Because adoption can occur through either licensing or infringement, we obtain the following result.

Proposition 3. *The probability of technology adoption is increasing in technological proximity θ.*

The probability of infringement (i.e. spillovers) is given by

$$\Pr(\text{infringement}) = \max\{0, F(x^*(\theta)) - F(\widetilde{x}(\theta))\}. \tag{9}$$

If $x^*(\theta) < \widetilde{x}(\theta)$ or proximity is within the range where $x^*(\theta) = \widetilde{x}(\theta)$, infringement does not occur so that small changes in proximity do not affect the probability of infringement. If $x^*(\theta) > \widetilde{x}(\theta)$, infringement can occur but the effects of changes in proximity on the likelihood of infringement are indeterminate.[8]

3 Data

We create a firm pairing dataset that tracks interactions in the market for technology between publicly held companies in North America with at least one patent. The centerpiece of the dataset is a newly created database on interactions in the market for technology. This database on interactions is complemented with firm level information from Compustat and patent data from the USPTO Patent Assignment Dataset (USPTO PAD, see Marco *et al.* , 2015).[9] In what follows we first describe the main components of our dataset as well as the match between each one of them. Next we describe the data and variables used in the regressions.

[8] The effects of proximity on the likelihood of infringement equal $\frac{\partial \Pr(\text{infringement})}{\partial \theta} = f(\widetilde{x}(\theta)) \frac{t_\theta(\widetilde{x}(\theta),\theta)}{t_x(\widetilde{x}(\theta),\theta)} - f(x^*(\theta)) \frac{t_\theta(x^*(\theta),\theta)}{t_x(x^*(\theta),\theta)}$.

[9] The match between Compustat and the USPTO Patent Assignment Dataset, which required a substantial amount of work, is described in great detail in the companion document "Matching assignees and assignors in the USPTO Patent Assignment Dataset to Compustat firms".

3.1 Dataset construction

Interactions in the market for technology. We follow Arora & Gambardella (2010) in defining the market for technology broadly as explicit transactions involving a formal exchange of knowledge for money or additional knowledge. We consider four major forms of technology exchange: patent trades, licensing, cross-licensing, and R&D alliances. The first two involve exchanges of knowledge for money while the last two involve exchanges of knowledge and perhaps also balancing payments. Patent trades, licensing, and cross-licensing involve an exchange of existing technologies while R&D alliances involve an exchange of future technologies (this might also be the case of ex-ante licensing agreements). Licensing and cross-licensing may be embedded in technological alliances of some sort.

Our main goal is to collect a list of interactions in the market for technology with broad coverage over time and across forms of exchange, sectors and technologies.[10] It is well understood that the market for technology is inherently opaque due to the enterprises' desire to keep their business strategies secret. To construct our database, we rely on voluntarily recorded patent assignments at the USPTO and compulsory disclosures to the SEC.[11] This implies that our data is a selected subset of the whole population of technology exchanges. It is unclear how severely affected by selection patent assignment records at the USPTO are. However, it is clear that SEC disclosures by definition offer a broader coverage of economically significant deals in which at least one the parties in the agreement (and often all of them) are publicly traded firms.[12] In what follows, we describe the construction of the datasets on each one of the different forms of technology transfer.

- *Patent trades:* We construct the dataset on patent trades between firms from the USPTO

[10] This implies that we intentionally discard well known databases specializing on specific sectors (mostly Pharma and Biotech) such as Thomson Reuter's Recap.

[11] SEC filings might also include voluntary disclosures. Our database may also include disclosures to analogous regulatory agencies from outside the US.

[12] Public companies are required to disclose "material" transactions in their filings. A "material" event is any significant event that affects the company's financial standing, such as a lawsuit, merger, employment of key personnel, joint venture, or license agreement. Public companies can be exempt from filing the standard SEC forms if they have fewer than 500 stockholders and less than $10 million in total assets.

PAD. This dataset provides detailed information on the changes in patent ownership.[13] Following Marco *et al.* (2015) and Serrano (2010) we define a transaction as an inter-firm patent trade if it is the second or subsequent transaction record for the patent, the conveyance text identifies the transaction as an "assignment of the assignor's interest" and neither the assignor nor the assignee are individuals according to the USPTO assignee identifiers.[14] This definition yields 295,068 assignments involving 797,211 patents (some of which are reassigned more than once) and 149,752 firms (after name harmonization and disambiguation).

- *Licensing:* We rely on ktMINE's Licensing Database and Thomson Reuters' Joint Venture & Strategic Alliances Database to obtain information on licensing.[15] Both of these databases apply publicly disclosed information (e.g. SEC filings). The main difference between the two datasets is that ktMINE collects all types of licensing deals while Thomson Reuters restricts attention to licensing within strategic alliances. These are the main characteristics of the two datasets:

  - ktMINE's Licensing Database includes over 12,000 licensing deals mostly extracted from SEC filings with filing dates on or after 1990. For each deal, we have detailed information on the identity of the licensor(s) and the licensee(s). We cleaned and harmonized licensor and licensee names obtaining 12,304 unique names.

  - Thomson Reuters Joint Venture & Strategic Alliances Database includes 14,270

[13] We work with a version of the USPTO PAD that covers 5,534,135 transactions recorded at the USPTO between January 1970 and January 2013 (inclusive). While the first transaction date is January 1970, the number of transactions recorded in the initial years is negligible. Data coverage seems sufficient for the years 1981-2012. Updated versions of the USPTO Patent Assignment Dataset can be found at: http://www.uspto.gov/learning-and-resources/electronic-data-products/patent-assignment-dataset.

[14] USPTO Assignee identifiers can be found in the PN_ASG_UPRD_69_13.TXT file of the PTMT Custom Extract 2013.

[15] Thomson Reuters' Joint Venture & Strategic Alliances Database was accessed through SDC Platinum. It tracks cooperative agreements by two or more separate entities that may (e.g. a joint venture) or may not (e.g. agreements involving licensing, research and development, cross technology transfer, etc.) result in the formation of a third entity. SDC offers a comprehensive coverage on the formation of all kinds of alliances by companies all over the world from 1988 on (earlier deals are covered on a much less comprehensive basis).

alliances with a licensing agreement flag. The identities of the parties in the licensing agreement are disclosed, but who the licensor(s) and the licensee(s) are is not specified. We carefully read a short memo available for each deal to find out who is who. Name cleaning and harmonization yielded 12,976 unique names.

- *Cross-licensing*: We rely on two datasets. The first one is a dataset that we have created on cross-licensing deals from disclosures to the SEC. The second one is a dataset on cross-technology transfer deals between firms in strategic alliances covered in the Thomson Reuters' Joint Venture & Strategic Alliances Database. Next we describe each one of these datasets.

    - The new database on cross-licensing deals was constructed as follows. First, we carried out an exhaustive search across all the SEC forms (about 22,500 forms) filed between 2000 and 2014 (both inclusive) containing the word "cross-licensing" (or related strings). We extracted information on the 4,375 instances in which the identity of the parties in the agreement was disclosed. Second, we searched in Google for cross-licensing deals by Compustat firms with a prominent patenting or R&D activity.[16] This second search resulted in 599 deals with the names of at least two cross-licensees. We appended the SEC and Google searches together and harmonized the names of the cross-licensees finding 2,608 unique names.
    - Thomson Reuters Joint Venture & Strategic Alliances Database includes 8,434 alliances with a cross-technology transfer agreement flag. As opposed to the cross-licensing deals compiled in the self-created dataset, these are transfers exclusively between firms collaborating in an alliance. We learnt from the short deal descriptions that the term "cross-technology transfer" not only refers to cross-licensing. Some deals involve research agreements or other unspecified forms of technology

[16] We define firms with a prominent patenting activity as firms with an average patent stock of more than 20 patents (according to the NBER-PDP match) and firms with a prominent R&D activity as firms that are not matched to patents but have average yearly R&D expenditures equal or greater than $3 million.

transfer. What remains clear is that whichever the exact technology transfer formula is it always involves bilateral exchanges of technology. Name cleaning and harmonization yielded 9,397 unique names.

- *R&D alliances*: We downloaded 16,160 alliances with an R&D agreement flag from Thomson Reuters Joint Venture & Strategic Alliances Database. R&D agreements involve the further development of a technology or its realization from scratch. What is exchanged in this case is future knowledge the creation of which would not be possible without the inputs or economic contribution by the partner. We cleaned and harmonized the names of the firms forming the joint venture obtaining 13,576 unique names.

It is important to understand some important features about the collected list of interactions. First, with most of the deals coming from SEC filings coverage is inevitably better for publicly traded firms.[17] Second, despite our efforts to amass the broadest possible amount of deals our data is far from being a census even for public firms.[18] That said, we believe that it is the most comprehensive dataset on interactions in the market for technology that can be put together for public firms. Finally, time coverage is acceptable for the post-1990 period.[19]

Compustat. Information in the Compustat Segment Dataset is used to define the product market of each firm. Additionally, several balance sheet items in Compustat North America Fundamentals (Annual) are used to control for firm specific attributes that might affect the probability of a match.

[17] Private firms do also show in deals disclosed to the SEC, but to a lesser extent. The filers of SEC filings are public firms. One of the parties of the deal is the filer itself (or a subsidiary). The other party can either be another public firm or a private one. So SEC deals are public to public or public to private.

[18] There are "non-material" deals not disclosed to the SEC or even "material" deals that are disclosed with redacted terms. For instance, LEXMARK INTERNATIONAL INC states in some of its SEC filings that it has in excess of one hundred patent cross-license agreements of which we were only able to find four. Similarly, STANDARD MICROSYSTEMS CORP states in its filings that it has patent cross-licensing agreements with more than thirty companies of which we found eight.

[19] This is necessarily the case for the licensing, cross-licensing and R&D alliances datasets which are exhaustive for the post-1990 period. The USPTO PAD covers the post-1980 period but reassignments remain relatively low until the mid-nineties (see Marco et al., 2015).

USPTO Patent Assignment Dataset (USPTO PAD). The USPTO PAD is used for three different purposes. The first one is to complement the dataset on interactions in the market for technology by tracking patent trades between firms. The second one is to measure the number of eventually granted utility patent applications per firm and year.[20] Such patent flow is constructed as the yearly number of original employer assignments of granted (by 2013) utility patents with application date on or after 1980.[21] The third one is to measure technological proximity between firms from the technology classes of their patents. To accomplish that we import technology classes from the BASIC_13 file in the USPTO PTMT Custom Extract 2013. The match between the USPTO PAD and Compustat is explained in detail in the companion document "Matching assignees and assignors in the USPTO Patent Assignment Dataset to Compustat firms".

Match between datasets. Firms in the dataset on interactions in the market for technology are linked to Compustat firms and USPTO PAD assignors/ees. In order to do that, we first produce the ASSIGNEE/OR-GVKEY file which links standardized and disambiguated assignor/ee names to Compustat GVKEYS.[22] Because the dataset on patent trades is constructed from the USPTO PAD, the ASSIGNEE/OR-GVKEY link already matches firms involved in patent assignments or reassignments to Compustat. The firms in the Licensing, Cross-licensing and R&D alliance datasets are matched to the ASSIGNEE/OR-GVKEY file

[20] The USPTO PAD is extensively described in Marco *et al.* (2015). We work with a version that covers 5,534,135 transactions recorded at the USPTO between January 1970 and January 2013 (inclusive). While the first transaction date is January 1970, the number of transactions recorded in the initial years is negligible. Data coverage seems sufficient for the years 1981-2012. Updated versions of the USPTO Patent Assignment Dataset can be found at: http://www.uspto.gov/learning-and-resources/electronic-data-products/patent-assignment-dataset.

[21] A transaction is defined as an employer assignment if it is the first transaction recorded for the patent, the execution date is prior to or on the grant date and the conveyance text identifies the transaction as an "assignment of the assignor's interest". This definition is consistent with the one proposed in Marco *et al.* (2015). Overall, 3,031,098 utility patents are originally assigned from inventors to corporations.

[22] The matching protocol, described in great detail in an independent file titled "Matching assignees and assignors in the USPTO Patent Assignment Dataset to Compustat firms", does the following: 1) standardizes and disambiguates assignor and assignee names in the USPTO Patent Assignment Dataset; 2) creates a dynamic corporate ownership tree with information from Compustat, SDC Platinum and Osiris; 3) matches the standardized and disambiguated assignee and assignor names to standardized firm names in the dynamic ownership tree and in the DYNASS file of the NBER Patent Data Project (see Hall *et al.* , 2001).

by name. Details on the match are provided in Appendix B.

Short description of the matched data. The final dataset includes 22,247 unique pairings matched to Compustat and the USPTO PAD. Most of the pairings interact through only one exchange mode, but up to 5,228 pairings interact through two or more forms of exchange. The dataset includes 4,707 unique firms (4,029 adopters and 3,833 providers with 3,155 firms doing both). The number of interactions per firm is highly skewed with most of the firms adopting from and/or providing to fewer than five firms. However, some firms have a remarkable number of interactions. The top-35 list of adopters and providers is provided in Table 1.

Execution dates are available for 22,114 pairings. Some pairings have multiple transaction records, with the average number of records per pairing being 7.9 and the total number of pairing-record observations 173,586. For patent trades, the number of transaction records is equivalent to the number of unique deals between firms. However, for the remaining modes of exchange the number of records generally exceeds the number of unique deals because some deals are collected multiple times through different SEC filings.[23] This implies that the dataset is better suited to capture the extensive margin (i.e. whether two firms ever interact) than the intensive margin (i.e. the number of interactions between firms).

Figure 1 shows the distribution of the number of pairings by execution year for every contractual mode of exchange. The number of interactions is almost negligible for the pre-1990 and post-2013 years. It is important to understand that part of the variation in the number of interactions over time is driven by sampling choices. For patent trades, sampling is stable over the whole period; for non-alliance related licensing and cross-licensing deals,

[23] Disambiguation of unique transactions is complex because the only information available for some deals is the name of the parties and the execution year. Admittedly, multiple entries between the same firms in a given execution year are likely to refer to the same deal (even though they could also be different deals). A very crude form of disambiguation is to keep just one pairing by execution year (even if the pairing exchanges technologies through different exchange modes). That results in 30,753 unique pairing-execution year observations. This number is substantially lower than the total number of records and not much higher than the total number of interactions.

sampling is relatively stable for the post-1995 period;[24] for within-alliance deals, sampling seems targeted around 1995.[25]

3.2 Samples and variables used in the econometric regressions

3.2.1 Samples

In order to study the determinants of firm matching in the market for technology in a regression framework we create three datasets with bilateral interactions between firms.

Dataset A. This dataset includes interactions between every one of the firms in the dataset on interactions with non-missing information on the variables used in the regressions (which are described below). This includes $3,605$ firms and $3,605 \cdot (3,605-1) = 12,992,420$ pairings. All the years are collapsed with the dataset resulting in a cross-section of adopter-provider pairings ij (i being the adopter and j the provider). The dependent variables are a series of dummy variables with value one if the ij pairing ever interacts in the market for technology (several dummy variables are created, one for every contractual mode of exchange). The explanatory variables are defined as means over all the available years. More details on the construction of the variables are provided below.

Dataset B. This dataset exploits the temporal dimension of the data on interactions to define predetermined explanatory variables. The dependent variables are defined as dummy variables with value one if the pairing interacts at least once in the market for technology during the post-2000 (including 2000) period. The explanatory variables are defined as means over the pre-2000 (excluding 2000) period. Only firms which interact in the market for technology in the post-2000 period and with pre-2000 information on the explanatory

[24] Independent licensing and cross-licensing deals are sampled from SEC forms filed on or after 2000. SEC filings essentially provide information on active deals. Many licensing and cross-licensing deals have terms of +5 years so post-2000 filings provide a substantial amount of information for the 1995-2000 period too.

[25] The sampling criteron used by SDC platinum to create the data on alliance-related deals (i.e. licensing and cross-licensing deals and R&D alliances) is not explicit, so time consistency cannot be assumed. The histograms suggest an uneven coverage with a peak around 1995 and a sharp decay after 2000.

variables are preserved. This includes $1,814$ firms and $1,814 \cdot (1,814 - 1) = 3,288,782$ pairings. Notice that this dataset is still a cross-section of ij pairings, even if it exploits the time dimension to construct pre-determined variables.

Dataset C. This dataset is a longitudinal panel of time dimension $T = 3$ with the time unit t being five year periods starting in $1995 - 2000$ and ending in $2005 - 2010$. The dependent variables are defined as dummy variables with value one if the firm pairing interacts at least once in the market for technology during the corresponding five-year period. The explanatory variables are defined as means over the corresponding five-year period.[26] We preserve a balanced panel with all the observations with non-missing information on all the variables all the panel periods. This includes 799 firms, $799 \cdot (799 - 1) = 637,602$ pairings and $3 \cdot 637,602 = 1,912,806$ observations.

3.2.2 Variables

The variables used in the regressions are defined below. Their descriptive statistics are provided in Table 2 for each one of the datasets described above.

Match. Seven dependent variables are defined. The first one is a dummy variable with value one if firm i adopts knowledge from firm j at least once in any of the different contractual forms of exchange and value zero otherwise. The remaining six are analogous dummy variables specific to every mode of exchange (i.e. patent trades, licensing outside alliances, licensing within alliances, cross-licensing outside alliances, cross-licensing within alliances and R&D alliances).

Market proximity. Following Bloom *et al.* (2013), we use the Compustat Segment Dataset on each firm's sales, broken down into four digit industry codes, to calculate market

[26] The econometric specification applied on Dataset C lags the explanatory variables by one (five-year) period. Therefore, the explanatory variables are constructed with data for the panel periods $1990 - 1995$, $1995 - 2000$ and $2000 - 2005$. So, overall, four five-year intervals are used to define the dependent and explanatory variables even if the time dimension of the panel is $T = 3$.

proximity. We define the vector $S_i = (S_{i1}, S_{i2,...}, S_{i1100})$ where S_{ik} is the share of sales of firm i in the four digit industry k, which runs from 1 to 1100 (the number of $SIC4$ codes across which we observe the distribution of sales). The Jaffe (1986) product market closeness measure is calculated as the uncentered correlation between the respective sales vectors of the two firms: $SIC_{ij} = (S_i S_j')[(S_i S_i')^{1/2}(S_j S_j')^{1/2}]^{-1} \times 100$. This index ranges between zero (minimum closeness) and one hundred (maximum closeness) depending on the degree of overlap in market proximity between firm pairings.

Technological proximity. We use the average share of patents per firm in each technology class as our measure of technological activity. We define the vector $T_i = (T_{i1}, T_{i2,...}, T_{i420})$ where $T_{i\tau}$ is the share of patents of firm i in technology class τ, which runs from 1 to 420 (the number of USPTO technology classes covered by the patents in our sample). The Jaffe (1986) technology closeness measure is calculated as the uncentered correlation between the respective technology vectors of the two firms: $TEC_{ij} = (T_i T_j')[(T_i T_i')^{1/2}(T_j T_j')^{1/2}]^{-1} \times 100$. This index ranges between zero (minimum closeness) and one hundred (maximum closeness) depending on the degree of overlap in technology between firm pairings.

Geographical proximity. We use the geographical location (i.e. U.S. counties or foreign country) of all the inventors on every patent to pinpoint the geographical distribution of a firm's research. For each firm we define the vector $G_i = (G_{i1}, G_{i2,...}, G_{i2412})$ where $G_{i\tau}$ is the share of inventors of firm i in location g, which runs from 1 to $2,412$ (reflecting $2,412$ U.S. counties and 149 foreign countries across which we observe the distribution of patents). Geographical closeness between firms is calculated as $GEO_{ij} = (G_i G_j')[(G_i G_i')^{1/2}(G_j G_j')^{1/2}]^{-1} \times 100$. Again, this index ranges between zero (minimum closeness) and one hundred (maximum closeness) depending on the degree of overlap in geographical proximity between firm pairings.

Controls. We define a series of firm specific variables to control for adopter and provider specific attributes. Such firm-specific controls are the natural log of (one plus) average yearly R&D expenditures, the natural log of (one plus) the average number of eventually granted yearly patent applications and the natural log of (one plus) the average number of workers. Averages are taken over all the available Compustat years for Dataset A, for pre-2000 years for Dataset B and for the corresponding five-year period for Dataset C. We also use as firm-specific controls a full set of industry (at the SIC2 level), technology class (at the NBER36 level) and geographical location (at the U.S. state level) dummy variables plus a variable that counts the number of years during which the pairing overlaps in Compustat. Additionally, for Dataset B we define a dummy variable with value one if the pair interacts in the market for technology in the pre-2000 period to control for adopter-provider fixed effects.

4 Visual network analysis

In this section we describe the dataset on interactions in the market for technology through a network analysis. We provide a visual portray of the market for technology which emerges from the linkage and aggregation of multiple bilateral interactions. We search for common patterns of interaction between firms and analyze if firms cluster by sectors of activity, technological fields and geographical locations. The network analysis is performed on a subsample of $19,578$ unique pairings between $3,753$ firms with non-missing information on SIC2 codes (from Compustat), technological classes (from USPTO PAD patents) and inventor geographical location (also from USPTO PAD patents). This dataset collapses all the years so it essentially captures the extensive margin of the market for technology (i.e. whether two firms ever interact) during the period 1990-2013.[27]

Figure 2 describes the network of interactions in the market for technology. Each node

[27] Some interactions take place earlier than 1980 (mostly for patent trades) and later than 2013 (some alliance-related deals from SDC Platinum). However, as Figure 1 makes clear, the bulk of interactions takes place between 1990 and 2013. The intensive margin (how many times two firms interact) is disregarded in the collapsed dataset.

represents a firm (with node size being proportional to the number of interactions of the firm) and each edge represents an interaction between two firms. Nodes are arranged following the Fruchterman & Reingold (1991) algorithm. Firms that appear closer to each other in the graph are more strongly connected. Notice that edges only provide information on the extensive margin (whether two firms interact or not) but not on the intensive margin (how many times they interact). Therefore, “strongly connected” means not that two firms interact more often with each other but rather that they tend to interact with the same firms. The names of the top-10 participants in the market for technology (counting both provisions and adoptions) are reported on the corresponding nodes. The network of interactions is colored by sector of activity (two digit SIC codes in Panel A), main technological field (NBER PDP 36 group aggregation, see Hall *et al.* , 2001, in Panel B) and US State (Panel C).

Two big clusters coexist in the market for technology: the Pharma and the ICT clusters, on the left and right sides of Figure 2 respectively. Such clusters are well delineated by specific sectors of activity, technological fields and geographical locations. Some firms in the transportation industry bridge the two big clusters. Interestingly, the Pharma and the ICT clusters have remarkably different structures.

The Pharma cluster is composed by a very strong core supplemented by a periphery. The central position in the Pharma cluster is filled with firms in the sector of activity SIC2-28 (Chemicals and allied products) and the technology class NBER-31 (Drugs). The periphery is filled with firms in the sector of activity SIC2-38 (Measuring, analyzing and controlling instruments; Photographic, medical and optical goods; Watches and clocks) and the technology classes NBER-32 (Surgery and medical instruments), NBER-33 (Biotechnology) and NBER-19 (Miscellaneous-chemical). Geographically wise, firms in the Pharma cluster are scattered over multiple US states without revealing a strong location pattern.

By constrast, the ICT cluster is composed of equally prominent sectors and technologies. By sector of activity, the ICT cluster comprises firms in SIC2-35 (Industrial and commercial machinery and computer equipment), SIC2-36 (Electronic and other electrical equipment and components, except computers) and SIC2-73 (Business Services) with all three sectors sharing an equally important role. By technological area, the ICT cluster includes NBER-22 (Computer hardware and software) and NBER-21 (Communications). Geographically wise, most of the firms in the ICT cluster are located in California. The firms that bridge the two big clusters belong mostly to SIC2-37 (Transportation Equipment) and NBER-19 (Miscellaneous-chemical).

Figure 3 provides interaction matrices describing the structure of interactions by sector of activity (Panel A), technology areas (Panel B) and US States (Panel C). Interaction matrices are helpful to intuitively understand the direction of knowledge flows between units, which is difficult to grasp from the interactions network. Each cell represents an interaction between two units with adopters being displayed in the y-axis and providers in the x-axis. Cell color intensity is increasing in the percentage of firm interactions taking place within each cell out of the total number of interactions. Three common patterns emerge from the interaction matrices. First, the diagonal sticks out meaning that within unit interactions are frequent. Second, certain rows and columns are colored all the way meaning that there are some prominent adopters and providers which adopt from and provide to a wide arrange of units. Therefore, not only within but also between unit exchanges are frequent with specific units being responsible for such cross-unit transfers. Finally, the most intensely colored rows correspond to units that also have intensely colored columns, meaning that frequent adopters also are frequent providers. This is not surprising given that certain forms of exchange such as cross-licensing and R&D alliances are bidirectional.

Beyond these common patterns, the interaction matrices reveal certain patterns by sector, technology and geography. Panel A reveals that some sectors of activity are responsible for most of the exchanges. These are the same salient sectors in the interactions network: SIC2-28 (Chemicals and allied products), SIC2-35 (Industrial and commercial machinery and computer equipment), SIC2-36 (Electronic and other electrical equipment and components, except computers), SIC2-37 (Transportation Equipment), SIC2-38 (Measuring, analyzing and controlling instruments; Photographic, medical and optical goods; Watches and clocks), SIC2-48 (Communications) and SIC2-73 (Business Services). Panel B reveals that the most prominent technology classes in the market for technology are also the ones with presence in the interactions network: NBER-19 (Miscellaneous-chemical), NBER-21 (Communications), NBER-22 (Computer hardware and software), NBER-31 (Drugs), NBER-32 (Surgery and medical instruments), NBER-33 (Biotechnology), NBER-45 (Power Systems) and NBER-46 (Semiconductor Devices). Finally, Panel C reveals that most of the firms interacting in the market for technologies are located in the expected states with big cities also with presence in the interactions network: California, Illinois, Massachusetts, New Jersey, New York and Texas.

It is interesting to compare Panels A and B in Figure 3 to similar figures in related papers that track knowledge flows through estimated interactions or patent citations. For instance, Panel A is comparable to Figure 1 in Manresa (2016), which summarizes firm-to-firm R&D spillovers estimated as parameters in a production function framework using the NBER-Compustat match in Bloom *et al.* (2013). The main providers of spillovers in Manresa (2016) are found to be SIC2-28, SIC2-36 and SIC2-48, which also happen to be some of the most salient provider industries in the market for technology in our analysis. This correspondence suggests that at least some of the spillovers documented by the regression-based spillover literature are internalized through prices. Panel B is analogous to Figure 1 in Acemoglu *et al.* (2016), an interaction matrix of cross-class citations between USPTO patents. The diagonal of the interaction matrix in Acemoglu *et al.* (2016) is much strongly colored than the rest, suggesting that patent citations take place mainly within technology field (defined as the NBER PDP 36 group aggregation, see Hall *et al.* , 2001). That stands in contrast with Panel B in Figure 3 where some columns and rows are at least as salient as the diagonal. It seems that knowledge diffusion might be more transversal between innovators (corporations) than between inventors.

5 Econometric analysis

The descriptive analysis shows that market, technological and geographical proximity help to explain the network of interactions in the market for technology. However, this positive association could be partly or entirely due to counfounding factors. We are particularly concerned about the following four identification problems. First, the three proximity metrics are likely to be correlated with each other because firms in the same industries use specific technologies and cluster in certain geographical locations. Simultaneously accounting for them all is necessary to ascertain their individual contribution to the probability of a match.

Second, simultaneity is not properly addressed in the visual analysis which relies on collapsed data that ignores the timing of the interactions. Collapsed data do not allow to tell whether firms interact because they are close or they are close because they interact. For instance, a firm may enter the market of the licensor upon having licensed its technology. We are interested in establishing direction from proximity in the respective metrics to interaction in the market for technology.

Third, firms are likely to have certain characteristics that make them more or less prone to participate in the market for technology (e.g. different patenting propensities). Such firm specific attributes need to be controlled for. Fourth, adopter-provider pairings are potentially affected by common shocks which may prompt them to both co-locate in similar spaces and interact more in the market for technology. This is indeed the reflection problem discussed in Manski (1993) which constitutes the central identification challenge in the R&D spillovers literature (see Bloom *et al.* , 2013). It is important to empirically distinguish proximity from technological shocks.

In practice, we face a trade-off between preserving the maximum number of observations and providing satisfactory solutions to the identification challenges discussed above. We propose the following three econometric specifications applied on each one of the datasets A to C described above which gradually move from preserving the maximum number of observations at the expense of offereing poorer identification to improving identification at the expense of giving up on an important number of observations.

Specification 1 - Collapsed matching and distance metrics, Dataset A. The following econometric specification provides a starting point to tease out the effects of each one of the distance metrics on the probability of a match,

$$m_{ij} = \beta_1 SIC_{ij} + \beta_2 TEC_{ij} + \beta_3 GEO_{ij} + \beta_4 X_i + \beta_5 X_j + \varepsilon_{ij} \tag{10}$$

where i and j index the adopter and the provider respectively; m_{ij} is a dummy variable with value one if i and j match in the market for technology at some point; SIC_{ij}, TEC_{ij}, and GEO_{ij} are the proximity metrics; X_i and X_j embody a full set of adopter and provider specific control variables respectively which include average yearly R&D expenditures, patent applications and number of employees plus a full set of sector of activity (at the SIC-2 level), technology class (at the NBER36 level) and geographic location (at the U.S. state level) dummy variables;[28] and ε_{ij} is an idiosyncratic error term.

Specification 2 - Predetermined distance metrics, Dataset B. In order to deal with simultaneity we split the sample into two periods (pre and post 2000) and construct predetermined explanatory variables. This results in the following specification

$$m_{ij\tau} = \beta_1 SIC_{ij\tau-1} + \beta_2 TEC_{ij\tau-1} + \beta_3 GEO_{ij\tau-1} + \beta_4 X_{i\tau-1} + \beta_5 X_{j\tau-1} + \beta_6 m_{ij\tau-1} + \varepsilon_{ij\tau} \tag{11}$$

where τ and $\tau - 1$ denote pre and post 2000 years respectively. Pre-2000 interactions $m_{ij\tau-1}$ are used to control for adopter-provider fixed effects.

Specification 3 - Adopter-provider fixed effects, Dataset C. In order to better control for adopter-provider shocks we exploit the longitudinal structure of Dataset C with

[28] Averages are taken over all the available Compustat years from 1980 onwards.

the following specification

$$m_{ijt} = \beta_1 SIC_{ijt-1} + \beta_2 TEC_{ijt-1} + \beta_3 GEO_{ijt-1} + \beta_4 X_{it-1} + \beta_5 X_{jt-1} + \phi_{ij} + \phi_t + \varepsilon_{ijt} \quad (12)$$

where t indexes five-year periods (starting in $1995 - 2000$ and ending in $2005 - 2010$), ϕ_{ij} are adopter-provider fixed effects and ϕ_t are time fixed effects. The explanatory variables are all lagged by one period as we are still concerned about simultaneity. Time-variant adopter and provider attributes are included in X_{it-1} and X_{jt-1}. This specification contols for adopter-provider fixed effects, but adopter-provider transitory shocks could still be in place. In order to entirely rule out a bias stemming from transitory shocks we would need instruments explaining exogenous variation in the proximity metrics over time.

6 Results

We present two sets of results. In a first stage, we use between and within adopter-provider variation to study whether the probability of a match is orthogonal or increasing with respect to the distance metrics after accounting for the identification problems discussed in the previous section. In a second stage we use between adopter-provider variation to study if the probability of a match is monotonic in the distance metrics.

Orthogonality Tables 3 to 5 report OLS estimates of Specifications 1 to 3 based on between adopter-provider variation in the Datasets A to C. Table 6 reports adopter-provider fixed effects estimates of Specification 3 based on within adopter-provider variation in the Dataset C.[29]

Table 3 reports the results of estimating Specification 1 on Dataset A. The coefficients of the distance metrics are all positive and statistically significant. This is also the case for the control variables, all of which enter the regressions with a positive sign except for

[29] The main reason for using linear rather than non-linear estimators (such as probit or logit) is that linear estimators allow dealing with fixed effects in a much simpler fashion through demeaning.

average yearly R&D in the patent trades equation in Column (2). These results suggest that the gains from trade are greater for firms that are closer in the market, technological and geographical dimensions. The coefficients of the explanatory variables suggest that larger firms, with higher average R&D expenditures and yearly patent applications are more likely to match in the market for technology.

Table 4 reports the results of estimating Specification 2 on Dataset B. The results are very similar to the ones in Table 4, which suggests that simultaneity is not a big issue in Specification 1. The coefficients of the distance metrics are positively signed and statistically significant, except for geographical proximity in the licensing within alliances equation in Column (4) which loses its significance. Also in Column (4), the coefficient of the provider's average yearly R&D is insignificant and the coefficient on the provider's average number of yearly employees is negative. The coefficients on the adopter's and provider's average yearly R&D remain insignificant in the patent trades equation in Column (2). The pre-2000 market interaction control is significant with a large positive coefficient in all the regressions. This suggests that the pre-2000 interactions at least partly control for adopter-provider fixed effects.

Table 5 reports pooled OLS estimates of Specification 3 on Dataset C. Notice that the longitudinal structure of Dataset C is not yet fully exploited here. The estimates based on the between adopter-provider variation in Dataset C are qualitatively very similar to the ones obtained with Datasets A and B. This implies that Dataset C is not fundamentally different from Datasets A and B, with all the datasets producing similar results when between adopter-provider variation is exploited.

Table 6 reports adopter-provider fixed effects estimates of Specification 3 on Dataset C. First, we focus on the results for non-alliance related interactions in Columns (2), (3) and (5).[30] The coefficient on market proximity is no longer significant and even turns negative

[30] In Section 3.1 we have shown that sampling is stable in the post-1995 period for patent trades and non-alliance related licensing and cross-licensing. This is sufficient to reliably exploit within variation in the data because the dependent variables in Dataset C are defined over the post-1995 period.

in the patent trades and licensing equations in Columns (2) and (3). The coefficient on technological proximity remains positive and significant for patent trades and cross-licensing in Columns (2) and (5). The coefficient on geographical proximity remains significant, albeit only at a 10%, in the patent trades equation in Column (2). Regarding the controls, the average number of patent applications of the adopter and the provider has a positive effect in the patent trades and cross-licensing equations in Columns (2) and (5). The number of employees of the provider also has a positive effect in the patent trades equation in Column (2). Quite surprisingly, the average R&D expenditure of the adopter and the provider have a negative effect in the licensing equation in Column (2).

Next, we discuss the results for within-alliance interactions in Columns (4), (6) and (7) of Table 6. Notice that potential unstable sampling limits our ability to safely exploit within variation.[31] For instance, the negative sign on most of the coefficients could be artificially driven by the decline in sampling intensity after the first panel period.[32] With this caveat in mind, we offer an inevitably more speculative interpretation of the results. Like in the non-alliance related interactions, the coefficient on market proximity is no longer significant except for the licensing equation in Column (4) where it is negative. The coefficients on technological and geographical proximity are negative in all the equations. This negative effect could mean that firms self-select into alliances when they are distant and do not have the ability to understand each other's technologies without the active involvement of the other party.

Interestingly, average R&D expenditure enters the regressions with a positive sign while the average number of patents enters the regressions with a negative sign. This is in contrast with the results for non-alliance related deals where the average number of patents mainly has a positive effect on the probability of a match. Again, one possible interpretation is

[31] In Section 3.1 we have shown that sampling seems targeted around 1995 for within-alliance deals, with the number of interactions declining substantially in the post-2000 period.

[32] A decline in sampling causes the dependent variables to systematically decline over time (i.e. transition from one in 1995-2000 to zero in 2000-2005 and 2005-2010) even when the explanatory variables increase, which translates into negative coefficients.

self-selection with firms which protect their technologies with patents engaging in arm's length transactions and firms which do not have well deliniated property rights over their technologies resorting to more complex contractual forms of exchange such as alliances and joint ventures.

The following summarizes the main findings in this section. First, between adopter-provider variation yields positive effects for all the proximity metrics and most of the controls, with the unanimously positive coefficients hinting a possible upward bias stemming from omitted adopter-provider shocks. Second, within adopter-provider variation yields less unanimous results. Technological and geographical proximity have a positive effect on the probability of a match for non-alliance related interactions, but a negative effect for interactions within alliances. Moreover, market proximity is no longer significant.

Monotonicity. Only between adopter-provider variation is used to study if the probability of a match is monotonically increasing in the distance metrics. This is mainly because we intend to produce predicted probabilities bounded between zero and one.[33] Specification 1 is (re)estimated on Dataset A by probit with the proximity metrics split into five dummy variables with value one if proximity is within a given interval (the base category is zero, the remaining groups are [1,20), [20,40), [40,60), [60,80) and [80,100]). Such decomposition allows us to flexibly estimate how the probability of a match varies for the different values of the distance metrics. The predicted probabilities resulting from the estimates are displayed in Figure 4.[34]

Figure 4 shows that the probability of a match is monotonically increasing in technological and geographical proximity for (almost) every single form of exchange. A different pattern emerges for market proximity with the probability of a match immediately raising above

[33] As mentioned above, nonlinear models (such as probit or logit) which bound predicted probabilities between zero and one are not well suited to accomodate fixed effects.

[34] Similar figures are obtained with between adopter-provider (probit) estimtaes of analogous dummy variables decompositions of Specification 2 (Dataset B) and Specification 3 (Dataset C). Within adopter-provider (linear) estimates of the dummy variables decomposition of Specification 3 (Dataset C) are monotonically increasing for technological proximity only (even though insignificant).

zero, but remaining fairly constant, for any positive value. The results for market proximity suggest a tension between the rent creation and rent dissipation effects of market proximity. This trade-off does not exist for technological and geographical proximity which have an unambiguously positive rent creation effect on the gains from technology adoption without triggering rent dissipation.

Finally, notice that Figure 4 allows us to gauge the magnitude of the effect of the proximity metrics on the probability of a match. For instance, for the dependent variable that groups all the modes of exchange ("All" in Row 1) and the technological proximity metric (TEC, in Column 2), the probability of a match between two firms that are perfectly close in the market and geographical spaces (i.e. SIC=100 and GEO=100) increases from 0 to 0.05 as technological proximity increases from 0 to 100.

7 Discussion

In this section we discuss the implications of the results presented in Section 6 for the R&D spillovers literature. There are two types of spillovers: rent spillovers and knowledge spillovers (Griliches, 1992; Hall *et al.* , 2010). The first type occurs when a firm purchases technologies at prices that do not reflect their usage value. The second type occurs when a firm's R&D is useful to another firm in doing its own research or when it spills over to other firms due to only partial excludability of knowledge. The topic of social returns to R&D is closely intertwined with that of knowledge spillovers.

A common approach to measuring pure knowledge spillovers consists in constructing the spillover pool available to firm i as $\sum_{j \neq i} \theta_{ij} G_j$,where θ_{ij} is a weighting matrix applied to the R&D stocks of other firms (G_j).[35] Coefficient estimates of the spillover pool (as an additional input in a production function framework) are assumed to capture pure knowledge spillovers. In other words, the positive effects of firm j's R&D on firm i's output are assumed not to be internalized by j. The wedge between the social and private rates of return to R&D (and

[35] Notice that the weights θ_{ij} are nothing but the proximity metrics that we have been using all along.

the socially optimal level of R&D) are calculated according to this implicit assumption of no internalization.

The extent to which the described approach is likely to deliver accurate estimates of pure spillovers depends on two important assumptions. First, the proximity weights θ_{ij} must be orthogonal to market transfers. Otherwise the external knowledge pool would include knowledge acquired on the market internalized by the provider. Second, adoption off the market of firm i from firm j must be monotonically increasing in proximity, or else the external knowledge pool would capture noise. Neither of these assumptions finds strong support in our theoretical and empirical results. Orthogonality is indeed rejected both theoretically and empirically. Monotonicity in the probability of a match has been predicted theoretically and corroborated empirically. However, monotonicity in the probability of adoption off the market cannot be tested empirically and has been questioned theoretically.[36]

It seems that estimates based on distance weighted spillover pools inevitably capture technology transfers partly internalized by technology providers. As a consequence, the wedge between the social and private rates of return to R&D might be narrower than typically estimated.

8 Conclusion

This paper offers an integral view of the market for technology by analyzing a newly created dataset on interactions between publicly held companies in North America that spans several years, contractual forms of exchange, industries and technologies. Special emphasis is placed on the relevance of market, technological and geographical proximity at shaping the market for ideas.

We apply a basic model of knowledge transfer between a technology provider and a

[36] Monotonicity is generally imposed as a natural by-product of the assumption that the gains from technology adoption are increasing in proximity. Our theoretical model shows that assuming the gains from technology adoption to be increasing in proximity is sufficient to generate monotonicity in overall adoption and in adoption through market transfers. However, it is not sufficient to generate monotonicity in the type of adoption that is conductive to pure spillovers (i.e. adoption through infringement).

technology adopter to guide the empirical analysis. A critical assumption of the model is that the gains from technology adoption are increasing in (market, technological or geographical) proximity. Adoption is allowed to take place either through a patent licensing agreement or, alternatively, through infringement. The model predicts that 1) the probability of adoption is increasing in market proximity, 2) the probability of adoption through a match in the market for technology is increasing in proximity and 3) the probability of adoption through infringement is not necessarily increasing in proximity.

The analysis has implications for the regression-based spillovers literature. In particular, the results suggest that the wedge between the social and private rates of return to R&D might be narrower than typically estimated.

A descriptive network analysis of the data on interactions shows that the interplay between market, technological and geographical proximity goes a long way at explaining the market for technology. A formal econometric analysis suggests that proximity has a positive effect on the probability of a match in the market for technology even after dealing with several identification issues. Estimates based on between adopter-provider variation yield positive effects for all the proximity metrics. Estimates based on within adopter-provider variation yield positive (negative) effects only for technological and geographical proximity for non-alliance (within-alliance) interactions. The probability of a match is monotonically increasing in technological and geographical proximity for (almost) every single form of exchange.

References

Acemoglu, Daron, Akcigit, Ufuk, & Kerr, William R. 2016. Innovation network. *Proceedings of the National Academy of Sciences*, **113**(41), 11483–11488.

Akcigit, Ufuk, Celik, Murat Alp, & Greenwood, Jeremy. 2016. Buy, Keep, or Sell: Economic Growth and the Market for Ideas. *Econometrica*, **84**(3), 943–984.

Ali, Ayfer, & Cockburn, Iain. 2016. *Buyer behavior in markets for technology: technology proximity between firm portfolio and in-licensed patents*. Mimeo.

Arora, A., Fosfuri, A., & Gambardella, A. 2001. *Markets for Technology: The Economics of Innovation and Corporate Strategy*. Cambridge, MA: MIT Press.

Arora, Ashish, & Fosfuri, Andrea. 2003. Licensing the market for technology. *Journal of Economic Behavior and Organization*, **52**(2), 277 – 295.

Arora, Ashish, & Gambardella, Alfonso. 2010. Ideas for rent: an overview of markets for technology. *Industrial and Corporate Change*, **19**(3), 775.

Arqué-Castells, Pere, & Spulber, Daniel F. 2017. *Technology Markets and R&D Spillovers*. Mimeo.

Bernstein, Jeffrey I., & Nadiri, M. Ishaq. 1988. Interindustry R&D Spillovers, Rates of Return, and Production in High-Tech Industries. *The American Economic Review*, **78**(2), 429–434.

Bloom, Nicholas, Schankerman, Mark, & Van Reenen, John. 2013. Identifying Technology Spillovers and Product Market Rivalry. *Econometrica*, **81**(4), 1347–1393.

Bottazzi, Laura, & Peri, Giovanni. 2003. Innovation and spillovers in regions: Evidence from European patent data. *European Economic Review*, **47**(4), 687 – 710.

Cassiman, Bruno, & Veugelers, Reinhilde. 2002. R&D Cooperation and Spillovers: Some Empirical Evidence from Belgium. *American Economic Review*, **92**(4), 1169–1184.

Cassiman, Bruno, & Veugelers, Reinhilde. 2006. In Search of Complementarity in Innovation Strategy: Internal R&D and External Knowledge Acquisition. *Management Science*, **52**(1), 68–82.

Ceccagnoli, Marco, Graham, Stuart J.H., Higgins, Matthew J., & Lee, Jeongsik. 2010. Productivity and the role of complementary assets in firms' demand for technology innovations. *Industrial and Corporate Change*, **19**(3), 839.

Cohen, Wesley M., & Levinthal, Daniel A. 1989. Innovation and Learning: The Two Faces of R&D. *The Economic Journal*, **99**(397), 569–596.

Drivas, Kyriakos, & Economidou, Claire. 2015. Is geographic nearness important for trading ideas? Evidence from the US. *The Journal of Technology Transfer*, **40**(4), 629–662.

Drivas, Kyriakos, Economidou, Claire, Karkalakos, Sotiris, & Tsionas, Efthymios G. 2016. Mobility of knowledge and local innovation activity. *European Economic Review*, **85**, 39 – 61.

Figueroa, Nicolas, & Serrano, Carlos J. 2013 (April). *Patent Trading Flows of Small and Large Firms*. Working Paper 18982. National Bureau of Economic Research.

Fosfuri, Andrea. 2006. The licensing dilemma: understanding the determinants of the rate of technology licensing. *Strategic Management Journal*, **27**(12), 1141–1158.

Fruchterman, Thomas M. J., & Reingold, Edward M. 1991. Graph drawing by force-directed placement. *Software: Practice and Experience*, **21**(11), 1129–1164.

Gambardella, Alfonso, & Giarratana, Marco S. 2013. General technological capabilities, product market fragmentation, and markets for technology. *Research Policy*, **42**(2), 315–325.

Gambardella, Alfonso, Giuri, Paola, & Luzzi, Alessandra. 2007. The market for patents in Europe. *Research Policy*, **36**(8), 1163 – 1183.

Gans, Joshua S., & Stern, Scott. 2010. Is there a market for ideas? *Industrial and Corporate Change*, **19**(3), 805.

Griliches, Zvi. 1992. The Search for R&D Spillovers. *The Scandinavian Journal of Economics*, **94**, S29–S47.

Hall, Bronwyn H., Jaffe, Adam B., & Trajtenberg, Manuel. 2001 (October). *The NBER Patent Citation Data File: Lessons, Insights and Methodological Tools*. Working Paper 8498. National Bureau of Economic Research.

Hall, Bronwyn H., Mairesse, Jacques, & Mohnen, Pierre. 2010. Chapter 24 - Measuring the Returns to R&D. *Pages 1033 – 1082 of:* Hall, Bronwyn H., & Rosenberg, Nathan (eds), *Handbook of the Economics of Innovation, Volume 2*. Handbook of the Economics of Innovation, vol. 2. North-Holland.

Jaffe, Adam B. 1986. Technological Opportunity and Spillovers of R&D: Evidence from Firms' Patents, Profits, and Market Value. *The American Economic Review*, **76**(5), 984–1001.

Lychagin, Sergey, Pinkse, Joris, Slade, Margaret E., & Reenen, John Van. 2016. Spillovers in Space: Does Geography Matter? *The Journal of Industrial Economics*, **64**(2), 295–335.

Manresa, Elena. 2016. *Estimating the Structure of Social Interactions Using Panel Data.* Mimeo.

Manski, Charles F. 1993. Identification of Endogenous Social Effects: The Reflection Problem. *The Review of Economic Studies*, **60**(3), 531–542.

Marco, Alan C., Myers, Amanda F., Graham, Stuart J.H., D'Agostino, Paul A, & Apple, Kirsten. 2015 (July). *The USPTO Patent Assignment Dataset: Descriptions and Analysis.* Working Paper 2015-2. USPTO.

Nash, John. 1953. Two-Person Cooperative Games. *Econometrica*, **21**(1), 128–140.

Rubinstein, Ariel. 1982. Perfect Equilibrium in a Bargaining Model. *Econometrica*, **50**(1), 97–109.

Serrano, Carlos J. 2010. The dynamics of the transfer and renewal of patents. *The RAND Journal of Economics*, **41**(4), 686–708.

Spulber, Daniel F. 2013. How Do Competitive Pressures Affect Incentives to Innovate When There Is a Market for Inventions? *Journal of Political Economy*, **121**(6), 1007–1054.

Spulber, Daniel F. 2015. How patents provide the foundation of the market for inventions. *Journal of Competition Law and Economics*, **11**(2), 271.

Spulber, Daniel F. 2016. Patent licensing and bargaining with innovative complements and substitutes. *Research in Economics*, **70**(4), 693 – 713. Special Issue on Industrial Organization.

Teece, David J. 1986. Profiting from technological innovation: Implications for integration, collaboration, licensing and public policy. *Research Policy*, **15**(6), 285 – 305.

Tables

Table 1. Top 35 adopters and providers

	Adopter / # firms it adopts from		Provider / # firms it provides to	
1	INTL BUSINESS MACHINES CORP	308	INTL BUSINESS MACHINES CORP	330
2	HP INC	264	HP INC	261
3	MICROSOFT CORP	214	MICROSOFT CORP	208
4	PFIZER INC	179	MOTOROLA SOLUTIONS INC	179
5	INTEL CORP	167	GENERAL ELECTRIC CO	163
6	MOTOROLA SOLUTIONS INC	166	INTEL CORP	163
7	ORACLE CORP	161	ORACLE CORP	163
8	SIEMENS AG	161	SIEMENS AG	162
9	GENERAL ELECTRIC CO	153	PFIZER INC	155
10	JOHNSON & JOHNSON	144	JOHNSON & JOHNSON	133
11	TOSHIBA CORP	129	DOW CHEMICAL	131
12	BRISTOL-MYERS SQUIBB CO	129	BRISTOL-MYERS SQUIBB CO	130
13	HITACHI LTD	128	ALCATEL-LUCENT	130
14	NEC CORP	125	LILLY (ELI) & CO	121
15	GLAXOSMITHKLINE PLC	122	TOSHIBA CORP	118
16	LILLY (ELI) & CO	120	HITACHI LTD	117
17	FUJITSU LTD	115	DU PONT (E I) DE NEMOURS	114
18	ALCATEL-LUCENT	113	NEC CORP	111
19	COMPAQ COMPUTER CORP	110	FUJITSU LTD	111
20	TEXAS INSTRUMENTS INC	109	GLAXOSMITHKLINE PLC	110
21	SONY CORP	108	TEXAS INSTRUMENTS INC	107
22	NOVARTIS AG	107	COMPAQ COMPUTER CORP	105
23	ABBOTT LABORATORIES	103	SUN MICROSYSTEMS INC	102
24	SUN MICROSYSTEMS INC	99	ABBOTT LABORATORIES	96
25	DU PONT (E I) DE NEMOURS	97	CISCO SYSTEMS INC	96
26	BAYER AG	97	NOVARTIS AG	96
27	CISCO SYSTEMS INC	96	BAYER AG	93
28	PANASONIC CORP	96	PANASONIC CORP	87
29	DOW CHEMICAL	94	MERCK & CO	84
30	APPLE INC	91	AT&T CORP	83
31	BASF SE	88	APPLE INC	83
32	SANOFI	80	HONEYWELL INTERNATIONAL INC	82
33	EASTMAN KODAK CO	80	BASF SE	82
34	LSI CORP	80	EASTMAN KODAK CO	81
35	DANAHER CORP	79	SONY CORP	81

Table 2. Descriptive statistics

Notes: R&D expenditure is in 2010 $million. The number of employees in thousands. The variables on interactions in the market for technology are dummy variables with value one if i adopts knowledge from j in the market in the corresponding contractual form of exchange.

	Dataset A					Dataset B					Dataset C				
	Mean	Median	S.D.	Min	Max	Mean	Median	S.D.	Min	Max	Mean	Median	S.D.	Min	Max
All market interactions ij	0.0015	0	0.0381	0	1	0.0015	0	0.0382	0	1	0.0023	0	0.0479	0	1
Patent purchases ij	0.0003	0	0.0173	0	1	0.0005	0	0.0217	0	1	0.0005	0	0.0225	0	1
Licensing, no alliance ij	0.0002	0	0.0155	0	1	0.0003	0	0.0172	0	1	0.0003	0	0.0179	0	1
Licensing, alliance ij	0.0003	0	0.0163	0	1	0.0001	0	0.01	0	1	0.0004	0	0.0191	0	1
Cross licensing, no alliance ij	0.0002	0	0.0143	0	1	0.0004	0	0.0196	0	1	0.0004	0	0.0207	0	1
Cross licensing, alliance ij	0.0003	0	0.018	0	1	0.0002	0	0.0126	0	1	0.0004	0	0.0191	0	1
R&D alliance ij	0.0006	0	0.0247	0	1	0.0003	0	0.0168	0	1	0.0009	0	0.03	0	1
Market proximity (SICij)	2.5	0	14.7	0	100	2.0	0	13.0	0	100	1.8	0	12.0	0	100
Technological proximity (TECij)	4.0	0	12.8	0	100	3.4	0	11.3	0	100	3.7	0	10.8	0	100
Geographical proximity (GEOij)	12.9	1	26.2	0	100	3.3	0	12.7	0	100	3.4	0	12.2	0	100
R&D_i	107.0	10.7	477.1	0	7,741	110.1	7.6	489.8	0	7,864	256.7	19.1	841.5	0	8,734
R&D_j	107.0	10.7	477.1	0	7,741	110.1	7.6	489.8	0	7,864	256.7	19.1	841.5	0	8,734
Patents_i	14.8	1.3	75.0	0	1,822	19.1	1.4	85.1	0	1,261	51.8	4.4	195.4	0	3,675
Patents_j	14.8	1.3	75.0	0	1,822	19.1	1.4	85.1	0	1,261	51.8	4.4	195.4	0	3,675
Employees_i	11.1	0.9	38.7	0	1,025	14.0	1.1	44.6	0	714	19.4	3.1	46.0	0	441
Employees_j	11.1	0.9	38.7	0	1,025	14.0	1.1	44.6	0	714	19.4	3.1	46.0	0	441
Overlap years in Compustat ij	10.6	9	9.6	0	66	18.0	16	10.7	2	66					
Market interactions pre-2000 ij						0.0	0	0.0	0	1					
Number of firms	3,605					1,814					799				
Number of pairings	12,992,420					3,288,782					637,602				
Number of observations	12,992,420					3,288,782					1,912,806				

Table 3. Results Specification 1

Notes: All the columns present OLS estimates. ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. Clustered (at the adopter-provider level) robust standard errors are reported in parentheses. The dependent variable is a dummy variable with value one if the pairing interacts in the market for technology (through the contractual mode of exchange specified on top of each column) and value zero otherwise. All the regressions include a full set of industry (at the SIC2 level), technology class (at the NBER PDP 36 group aggregation level), and geographical location (U.S. states or foreign country) fixed effects at the adopter and provider level. The number of overlap years between the adopter and the provider is included as a control.

	All	Patent trades	Licensing, no alliance	Licensing, alliance	Cross-licensing, no alliance	Cross-licensing, alliance	R&D alliance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Market proximity	0.000071***	0.000011***	0.000017***	0.000015***	0.000016***	0.000016***	0.000026***
	(0.000004)	(0.000001)	(0.000002)	(0.000002)	(0.000001)	(0.000001)	(0.000002)
Technological proximity	0.000202***	0.000041***	0.000044***	0.000038***	0.000039***	0.000044***	0.000083***
	(0.000009)	(0.000002)	(0.000003)	(0.000003)	(0.000003)	(0.000004)	(0.000006)
Geographical proximity	0.000045***	0.000018***	0.000008***	0.000005***	0.000007***	0.000010***	0.000017***
	(0.000003)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000001)	(0.000002)
ln(R&D_i)	0.000185***	0.000003	0.000056***	0.000038***	0.000021**	0.000036**	0.000113***
	(0.000040)	(0.000008)	(0.000011)	(0.000012)	(0.000008)	(0.000016)	(0.000025)
ln(R&D_j)	0.000179***	0.000008*	0.000048***	0.000030***	0.000021***	0.000036***	0.000113***
	(0.000015)	(0.000005)	(0.000005)	(0.000005)	(0.000004)	(0.000006)	(0.000009)
ln(patents_i)	0.001237***	0.000250***	0.000131***	0.000281***	0.000233***	0.000375***	0.000587***
	(0.000129)	(0.000017)	(0.000020)	(0.000041)	(0.000030)	(0.000058)	(0.000078)
ln(patents_j)	0.001259***	0.000287***	0.000151***	0.000252***	0.000232***	0.000375***	0.000587***
	(0.000057)	(0.000015)	(0.000014)	(0.000020)	(0.000018)	(0.000028)	(0.000037)
ln(employees_i)	0.000540***	0.000073***	0.000082***	0.000129***	0.000046***	0.000180***	0.000295***
	(0.000077)	(0.000012)	(0.000018)	(0.000024)	(0.000015)	(0.000031)	(0.000047)
ln(employees_j)	0.000524***	0.000136***	0.000053***	0.000048***	0.000046***	0.000180***	0.000295***
	(0.000023)	(0.000008)	(0.000007)	(0.000006)	(0.000006)	(0.000015)	(0.000018)
R^2	.015	.0035	.0032	.0033	.0035	.0049	.0077
Observations	12,992,420	12,992,420	12,992,420	12,992,420	12,992,420	12,992,420	12,992,420

Table 4. Results Specification 2

Notes: All the columns present OLS estimates. ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. Clustered (at the adopter-provider level) robust standard errors are reported in parentheses. The dependent variable is a dummy variable with value one if the pairing interacts in the market for technology (through the contractual mode of exchange specified on top of each column) and value zero otherwise. All the regressions include a full set of industry (at the SIC2 level), technology class (at the NBER PDP 36 group aggregation level), and geographical location (U.S. states or foreign country) fixed effects at the adopter and provider level. The number of overlap years between the adopter and the provider is included as a control.

	All	Patent trades	Licensing, no alliance	Licensing, alliance	Cross-licensing, no alliance	Cross-licensing, alliance	R&D alliance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Market proximity (SICij)	0.000071***	0.000018***	0.000018***	0.000005***	0.000027***	0.000006***	0.000012***
	(0.000004)	(0.000002)	(0.000002)	(0.000001)	(0.000003)	(0.000001)	(0.000002)
Technological proximity (TECij)	0.000199***	0.000064***	0.000049***	0.000012***	0.000066***	0.000015***	0.000038***
	(0.000005)	(0.000003)	(0.000003)	(0.000001)	(0.000003)	(0.000002)	(0.000002)
Geographical proximity (LOCij)	0.000041***	0.000025***	0.000007***	0.000000	0.000007***	0.000002**	0.000013***
	(0.000003)	(0.000002)	(0.000002)	(0.000001)	(0.000002)	(0.000001)	(0.000002)
ln(R&D_i)	0.000133***	-0.000002	0.000056***	0.000011**	0.000041***	0.000026***	0.000039***
	(0.000019)	(0.000012)	(0.000009)	(0.000006)	(0.000009)	(0.000007)	(0.000008)
ln(R&D_j)	0.000158***	0.000017	0.000053***	0.000006	0.000042***	0.000026***	0.000040***
	(0.000019)	(0.000011)	(0.000008)	(0.000005)	(0.000009)	(0.000007)	(0.000008)
ln(patents_i)	0.000822***	0.000282***	0.000089***	0.000057***	0.000237***	0.000122***	0.000170***
	(0.000034)	(0.000019)	(0.000014)	(0.000010)	(0.000018)	(0.000012)	(0.000014)
ln(patents_j)	0.000830***	0.000298***	0.000096***	0.000061***	0.000237***	0.000122***	0.000170***
	(0.000034)	(0.000020)	(0.000014)	(0.000009)	(0.000018)	(0.000012)	(0.000014)
ln(employees_i)	0.000290***	0.000079***	0.000066***	0.000039***	0.000052***	0.000018*	0.000093***
	(0.000027)	(0.000015)	(0.000012)	(0.000008)	(0.000014)	(0.000009)	(0.000011)
ln(employees_j)	0.000352***	0.000191***	0.000049***	-0.000012	0.000057***	0.000019**	0.000096***
	(0.000028)	(0.000017)	(0.000012)	(0.000007)	(0.000014)	(0.000009)	(0.000011)
Market interactions ij	0.148709***	0.038096***	0.036918***	0.012349***	0.057596***	0.017986***	0.040354***
	(0.004422)	(0.002423)	(0.002347)	(0.001375)	(0.002891)	(0.001660)	(0.002422)
R^2	.046	.011	.013	.0042	.024	.0062	.016
Observations	3,288,782	3,288,782	3,288,782	3,288,782	3,288,782	3,288,782	3,288,782

Table 5. Results Specification 3, pooled OLS

Notes: All the columns present OLS estimates. ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. Clustered (at the adopter-provider level) robust standard errors are reported in parentheses. The dependent variable is a dummy variable with value one if the pairing interacts in the market for technology (through the contractual mode of exchange specified on top of each column) and value zero otherwise. All the regressions include a full set of industry (at the SIC2 level), technology class (at the NBER PDP 36 group aggregation level), and geographical location (U.S. states or foreign country) fixed effects at the adopter and provider level. The number of overlap years between the adopter and the provider is included as a control.

	All	Patent trades	Licensing, no alliance	Licensing, alliance	Cross-licensing, no alliance	Cross-licensing, alliance	R&D alliance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Market proximity	0.000135***	0.000027***	0.000033***	0.000021***	0.000047***	0.000010***	0.000043***
	(0.000010)	(0.000004)	(0.000004)	(0.000004)	(0.000005)	(0.000003)	(0.000006)
Technological proximity	0.000434***	0.000084***	0.000072***	0.000060***	0.000111***	0.000068***	0.000173***
	(0.000012)	(0.000005)	(0.000005)	(0.000004)	(0.000007)	(0.000004)	(0.000007)
Geographical proximity	0.000103***	0.000036***	0.000016***	0.000000	0.000011***	0.000005***	0.000065***
	(0.000008)	(0.000004)	(0.000003)	(0.000002)	(0.000003)	(0.000002)	(0.000007)
ln(R&D_i)	0.000337***	-0.000001	0.000060***	0.000104***	0.000016	0.000094***	0.000250***
	(0.000030)	(0.000015)	(0.000010)	(0.000012)	(0.000011)	(0.000011)	(0.000018)
ln(R&D_j)	0.000335***	0.000026*	0.000057***	0.000074***	0.000016	0.000094***	0.000250***
	(0.000030)	(0.000015)	(0.000010)	(0.000012)	(0.000011)	(0.000011)	(0.000018)
ln(patents_i)	0.000553***	0.000167***	0.000030*	0.000004	0.000279***	0.000114***	-0.000036
	(0.000055)	(0.000023)	(0.000018)	(0.000019)	(0.000026)	(0.000024)	(0.000033)
ln(patents_j)	0.000572***	0.000153***	0.000054***	0.000049**	0.000279***	0.000114***	-0.000036
	(0.000056)	(0.000023)	(0.000019)	(0.000021)	(0.000026)	(0.000024)	(0.000033)
ln(employees_i)	0.001005***	0.000175***	0.000158***	0.000215***	0.000084***	0.000192***	0.000576***
	(0.000056)	(0.000026)	(0.000020)	(0.000020)	(0.000024)	(0.000021)	(0.000034)
ln(employees_j)	0.000980***	0.000255***	0.000067***	0.000148***	0.000084***	0.000192***	0.000576***
	(0.000056)	(0.000026)	(0.000019)	(0.000022)	(0.000024)	(0.000021)	(0.000034)
R^2	.026	.0048	.0053	.0043	.0084	.0058	.014
Observations	1,912,806	1,912,806	1,912,806	1,912,806	1,912,806	1,912,806	1,912,806

Table 6. Results Specification 3, adopter-provider fixed effects

Notes: All the columns present adopter-provider fixed effects estimates (OLS estimtaes on the demeaned variables). ***, ** and * indicate significance at a 1%, 5% and 10% level respectively. Clustered (at the adopter-provider level) robust standard errors are reported in parentheses. The dependent variable is a dummy variable with value one if the pairing interacts in the market for technology (through the contractual contractual mode of exchange specified on top of each column) and value zero otherwise. The number of overlap years between the adopter and the provider is not included as a control because Dataset C is defined as a balanced panel (i.e. all the firms remain in the dataset and overlap during all the years). The explanatory variables are lagged by one period.

	All	Patent trades	Licensing, no alliance	Licensing, alliance	Cross-licensing, no alliance	Cross-licensing, alliance	R&D alliance
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Market proximity	-0.000001	-0.000002	-0.000005	-0.000028***	0.000009	0.000006	0.000004
	(0.000021)	(0.000009)	(0.000008)	(0.000009)	(0.000011)	(0.000008)	(0.000014)
Technological proximity	-0.000036***	0.000010*	0.000001	-0.000032***	0.000013***	-0.000006*	-0.000055***
	(0.000010)	(0.000005)	(0.000005)	(0.000005)	(0.000004)	(0.000004)	(0.000006)
Geographical proximity	-0.000005	0.000008*	0.000004	-0.000012***	0.000002	-0.000005*	-0.000012***
	(0.000007)	(0.000004)	(0.000002)	(0.000003)	(0.000003)	(0.000003)	(0.000004)
ln(R&D_i)	0.000161**	0.000006	-0.000061***	0.000065**	0.000004	0.000105***	0.000143***
	(0.000065)	(0.000031)	(0.000020)	(0.000033)	(0.000025)	(0.000022)	(0.000041)
ln(R&D_j)	0.000045	-0.000005	-0.000078***	-0.000024	0.000004	0.000105***	0.000143***
	(0.000065)	(0.000035)	(0.000021)	(0.000030)	(0.000025)	(0.000022)	(0.000041)
ln(patents_i)	-0.000915***	0.000062*	-0.000001	-0.000437***	0.000082***	-0.000345***	-0.001022***
	(0.000084)	(0.000034)	(0.000028)	(0.000041)	(0.000031)	(0.000046)	(0.000063)
ln(patents_j)	-0.000820***	0.000125***	-0.000010	-0.000452***	0.000082***	-0.000345***	-0.001022***
	(0.000084)	(0.000034)	(0.000027)	(0.000045)	(0.000031)	(0.000046)	(0.000063)
ln(employees_i)	0.000185	-0.000015	-0.000059	-0.000035	0.000023	0.000181**	-0.000014
	(0.000149)	(0.000068)	(0.000045)	(0.000066)	(0.000058)	(0.000074)	(0.000107)
ln(employees_j)	0.000341**	0.000195***	-0.000054	-0.000225***	0.000023	0.000181**	-0.000014
	(0.000155)	(0.000073)	(0.000052)	(0.000074)	(0.000058)	(0.000074)	(0.000107)
R^2	.00076	.000051	.00014	.0013	.00019	.00057	.0021
Observations	1,912,806	1,912,806	1,912,806	1,912,806	1,912,806	1,912,806	1,912,806

Figures

Figure 1. Interactions by execution year

This figure provides histograms for the number of unique pairing interactions by execution year for each contractual mode of exchange. Only one pairing interaction per execution year is kept (i.e. each pairing is counted only once per year per contractual mode of exchange even if it has more than one record with the same execution date for the corresponding mode of exchange).

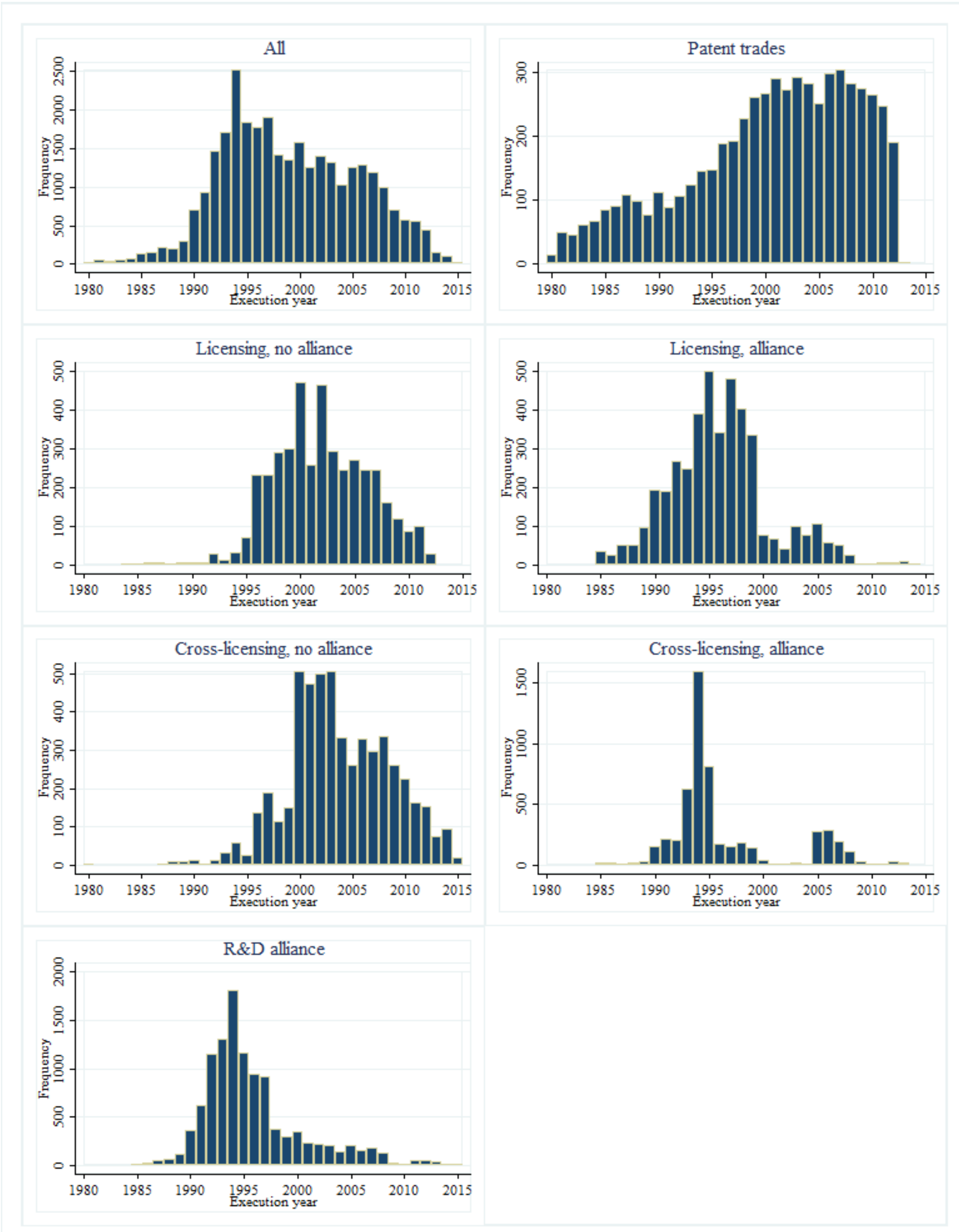
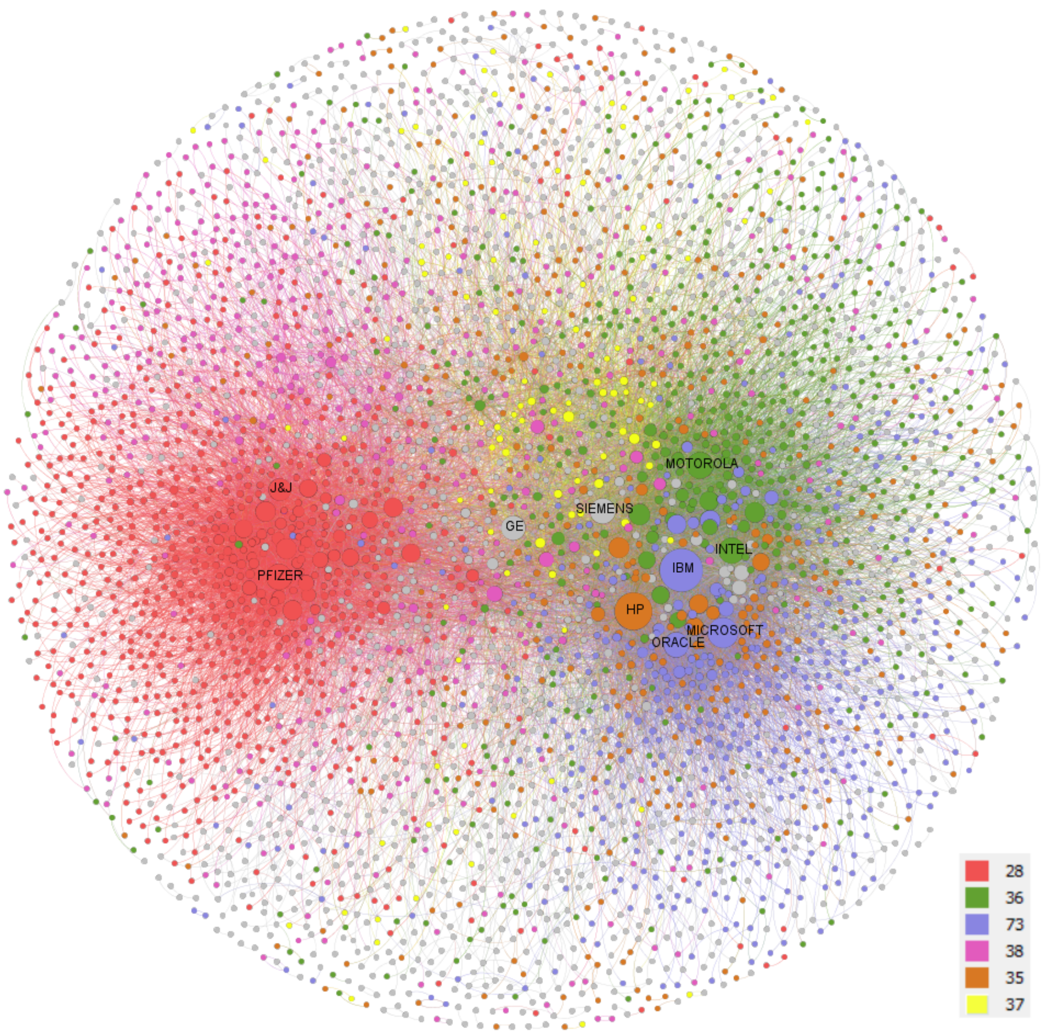


Figure 2. Network of interactions in the market for technology

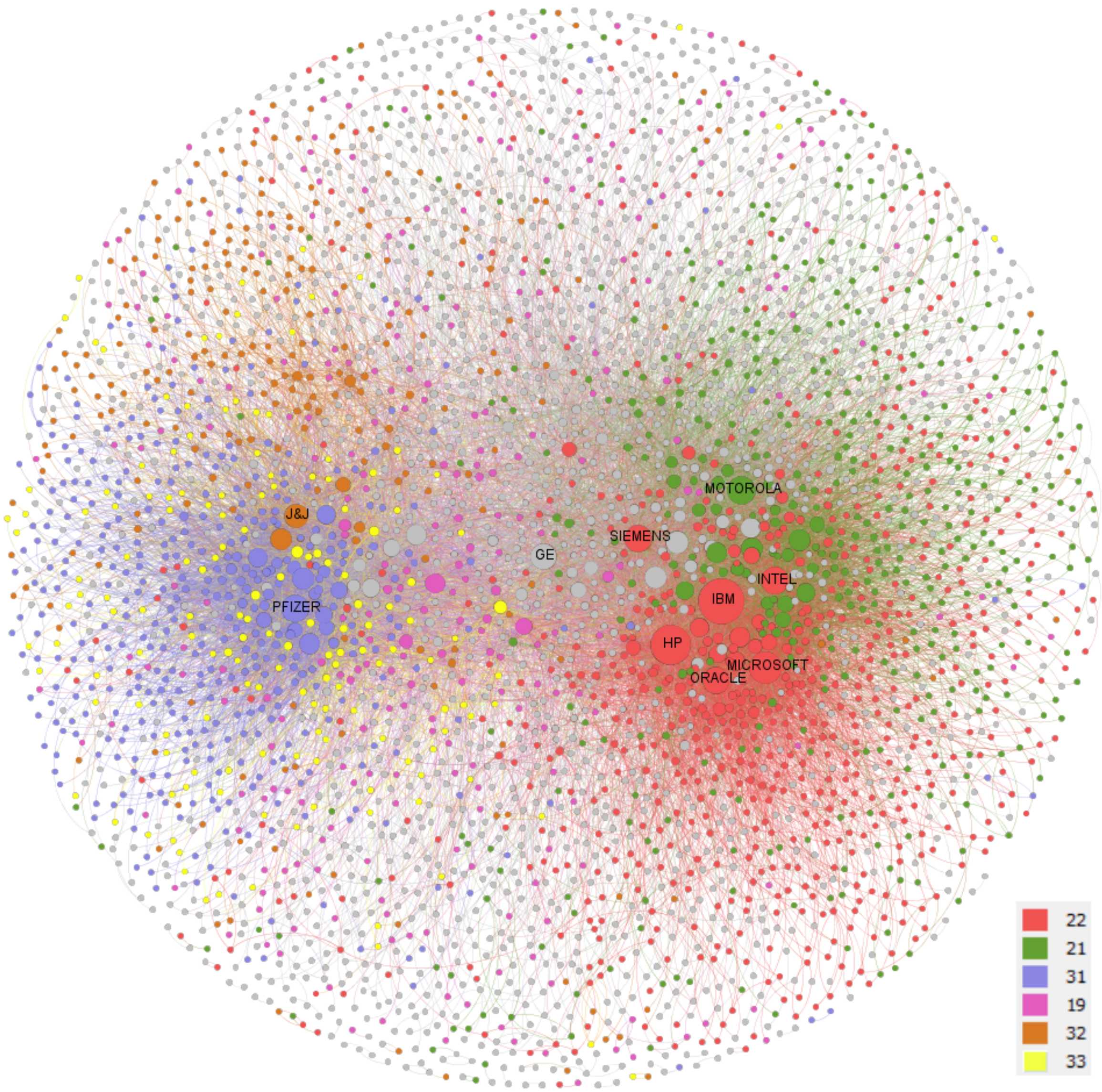
Each node represents a firm (with node size being proportional to the number of interactions of the firm) and each edge represents an interaction between two firms. Nodes are arranged following the Fruchterman and Reingold (1991) algorithm. The graphs have been generated with Gephi.

A. Colored by main sector of activity



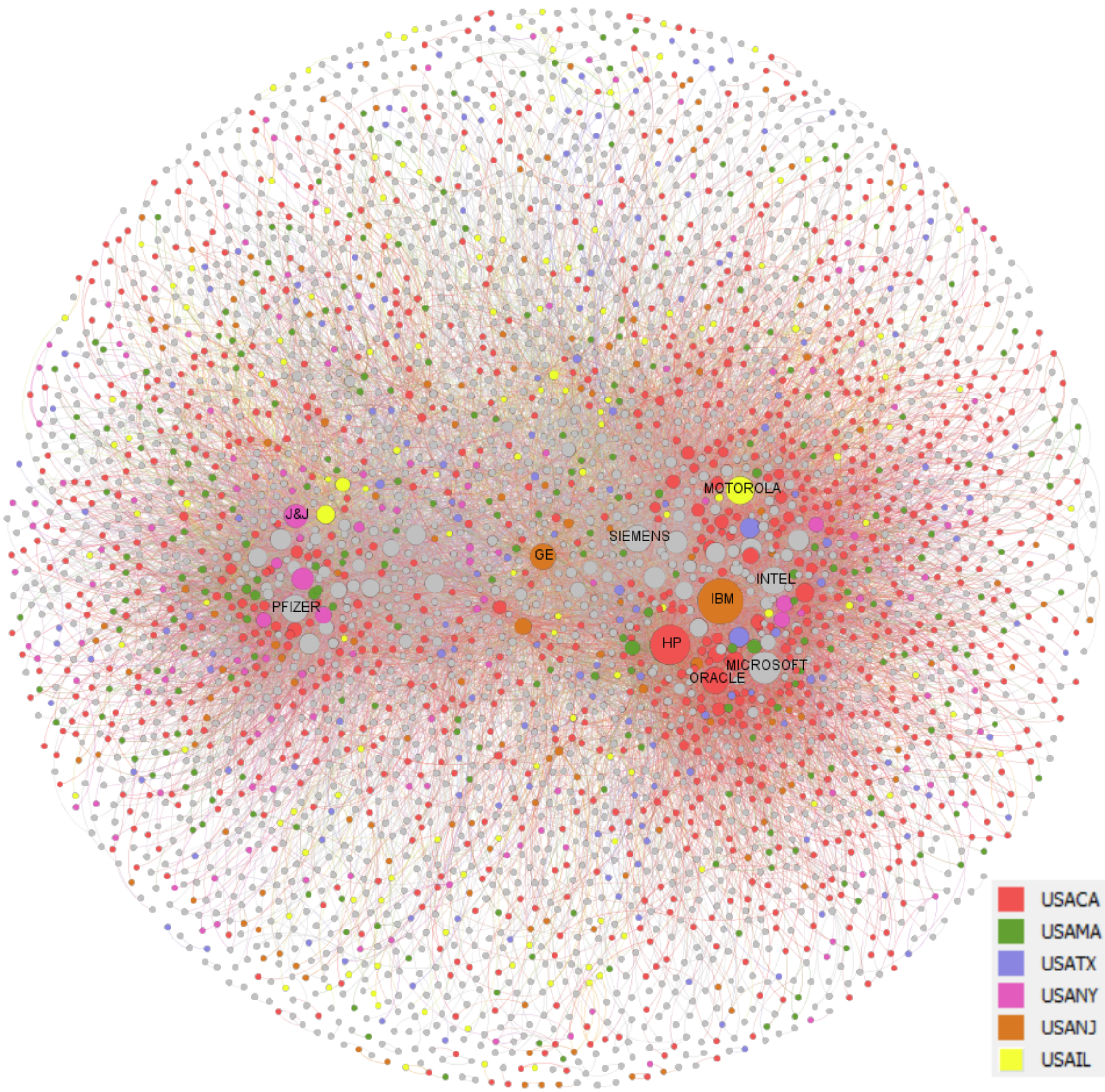
Notes: Each color represents a SIC-2 code. Only the following six sectors (with the highest number of interactions) are colored: 28-Chemicals and allied products; 36-Electronic and other electrical equipment and components (except computers); 73-Business Services; 38-Measuring, analyzing and controlling instruments; Photographic, medical, and optical goods; Watches and clocks; 35-Industrial and commercial machinery and computer equipment; and 37-Transportation Equipment.

B. Colored by main technological area



Notes: Each color represents a technology class from the 36 technology class aggregation in the NBER PDP (see Hall et al., 2001). The main technology class is defined as the modal class of the patents of the firm. Only the following six classes (with the highest number of interactions) are colored: 22-Computer hardware and software; 21-Communications; 31-Drugs; 19-Miscellaneous-chemical; 32-Surgery and medical instruments; and 33-Biotechnology.

C. Colored by main inventor location of the firm

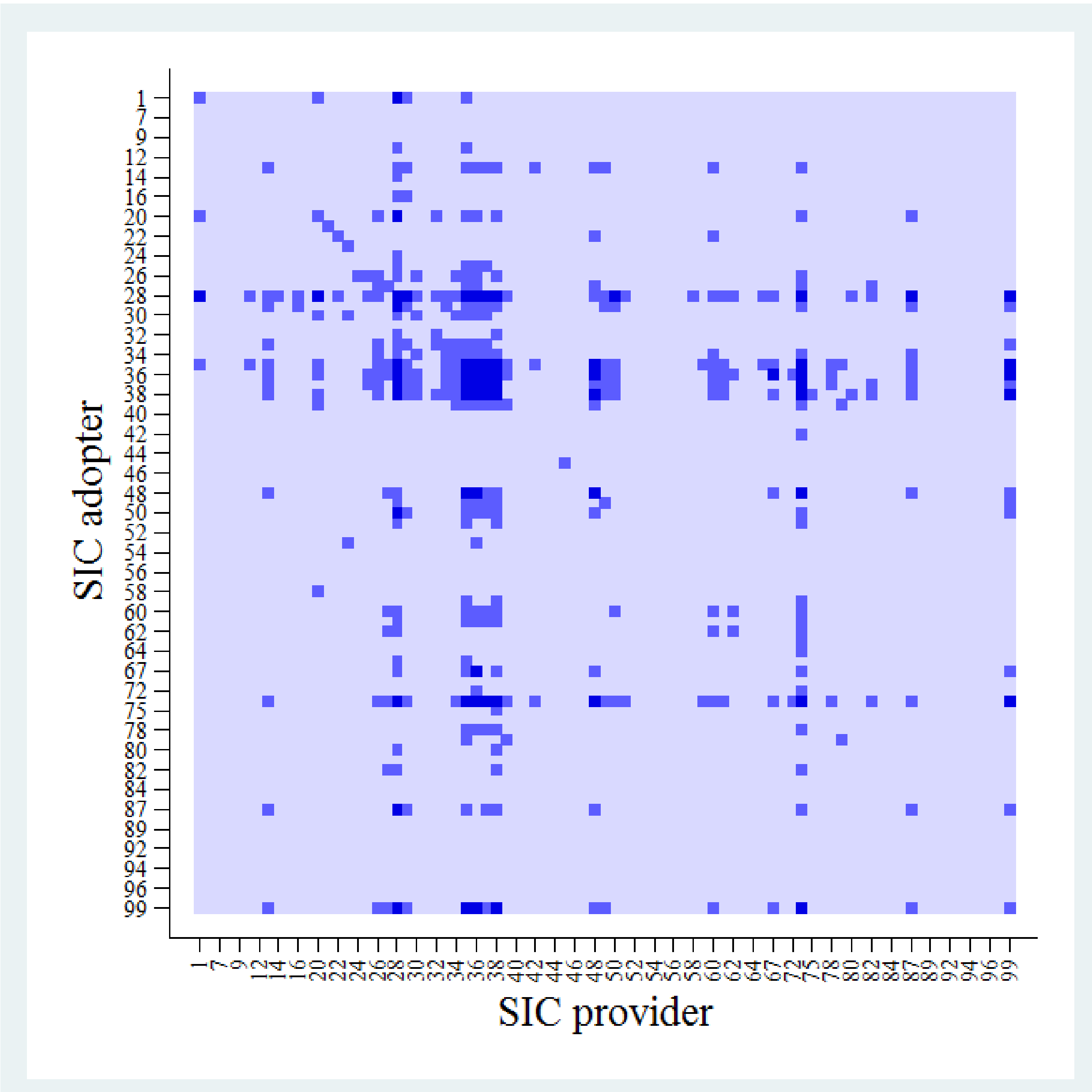


Notes: Each color represents a US State corresponding to the modal inventor location of the firm (which turns out to be the headquarter of the firm in most of the cases). The following six states (with the highest number of interactions) are colored: California (CA), Massachusetts (MA), Texas (TX), New York (NY), New Jersey (NJ) and Illinois (IL).

Figure 3. Interaction matrices

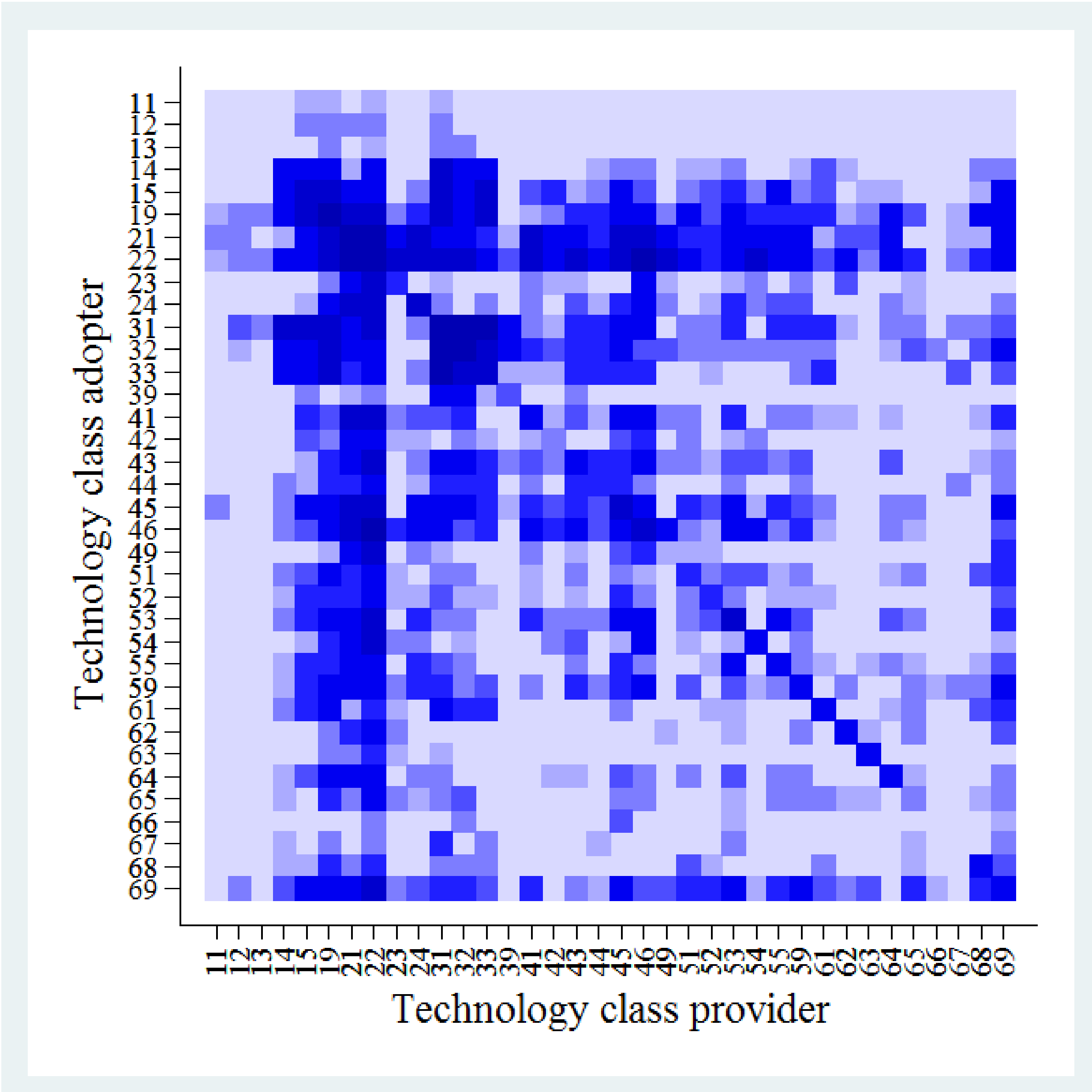
The graphs in Figure 2 show the structure of interactions aggregated by sector of activity, technology areas and US States. Each cell represents an interaction between two units (i.e. SIC2 sectors, NBER PDP technology classes or US States). Cell color intensity is increasing in the percentage of firm interactions taking place within each cell out of the total number of interactions.

A. Interactions by sector of activity



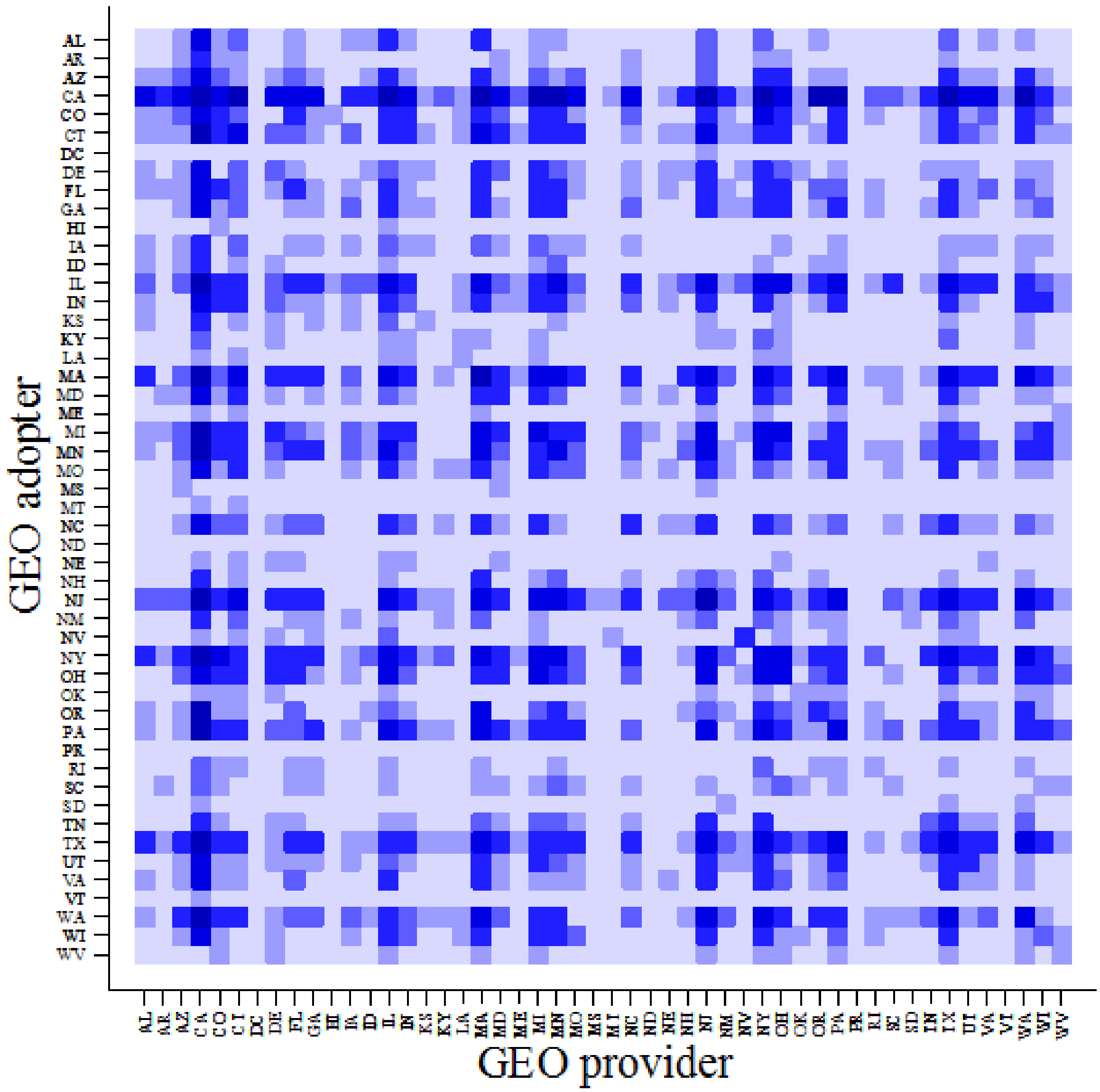
Notes: The main industry of the firm is difined as the SIC-2 code in Compustat.

B. Interactions by technology class



Notes: The main technology class is defined as the modal class of the patents of the firm.

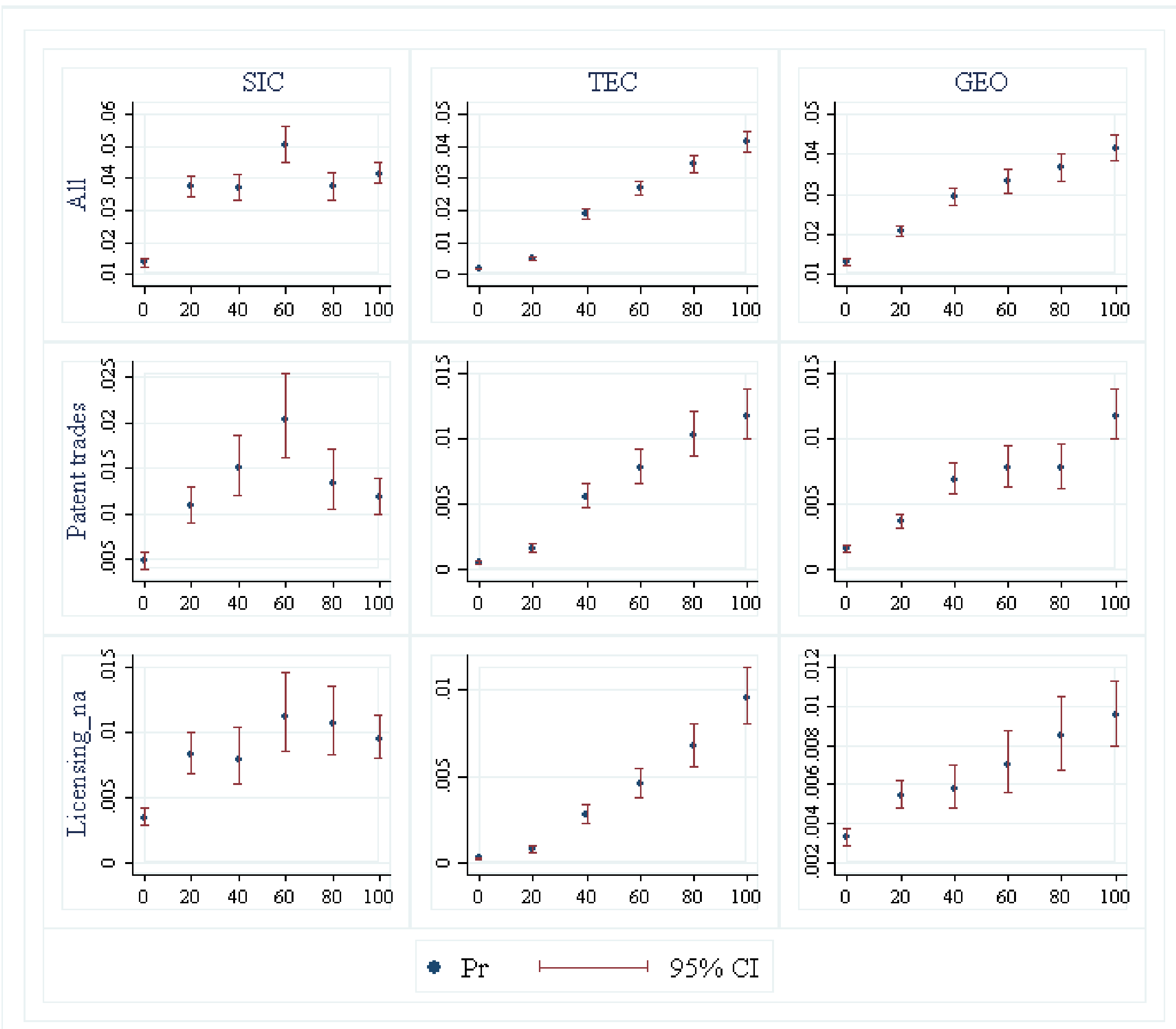
C. Interactions by location



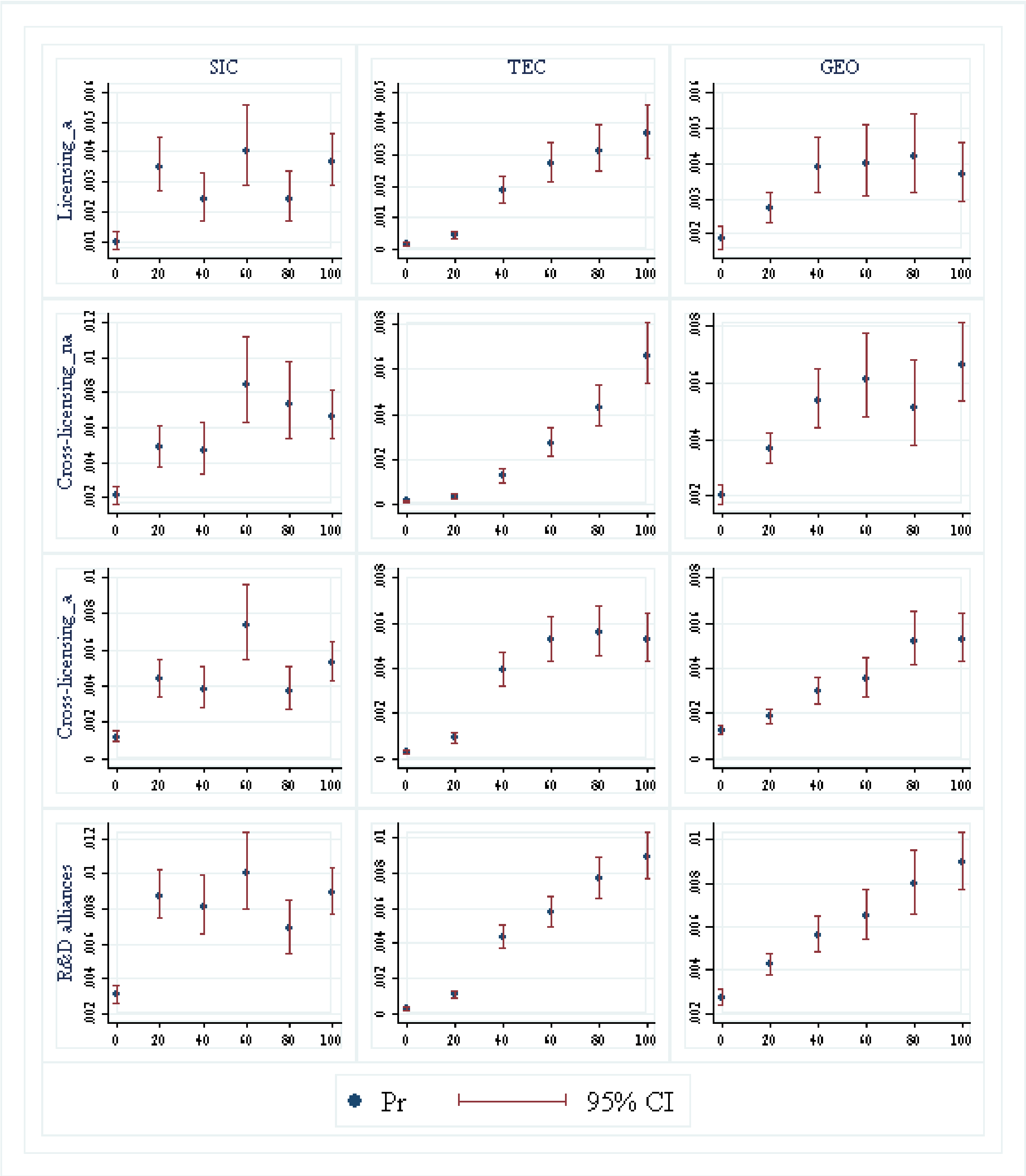
Notes: The main research location is defined as the U.S. state where most of the inventors of the firm's patents are located.

Figure 4. Predicted probability of a match

The predicted probability of a match (y-axis) is plotted against the proximity metrics (x-axis) using probit estimates from Specification 1 (Dataset A) but with the proximity metrics split into five dummy variables with value one if proximity is within the corresponding interval (the base category is zero, the remaining groups are [1,20), [20,40), [40,60), [60,80) and [80,100]). Rows are regression specific for the indicated dependent variable. Columns are distance metric specific. The explanatory variables are evaluated at their mean. The proximity metrics are set at their maximum value (100) when used as controls (e.g. TEC=100 and GEO=100 when calculating the probabilities for the different intervals of SIC in column one). The industry sector, technology class and geographic location dummy variables included in the regressions when estimating the results displayed in Tables 3 and 4 have been exluded from the probit regressions. This is because many of such dummy variables perfectly predict the outcomes and are authomatically dropped from the regressions, or do not generate standard errors when forced to remain in the probit regressions. So we have dropped the full set of dummy variables in order to estimate the probits on the samples used in Tables 3 and 4 and to produce predictions with confidence intervals. Including or excluding the full set of dummy variables barely affects the estimates.



(Continuation)



A Theory Appendix

Proof of Proposition 2. From the definition of $G(x,\theta)$ and the definition of the outside options $\pi^P(x,\theta)$ and $\pi^A(x,\theta)$, it follows that $G(x,\theta)$ is continuously differentiable in its arguments for all x except at $\widetilde{x}$ where there is a discontinuity. If $x^* < \widetilde{x}$, licensing occurs if $x^* \leq x < \widetilde{x}$ and the outside option is non-adoption.

$$G(x,\theta) = (1-\lambda)t(x,\theta) - c^P - c^A.$$

If $x^* > \widetilde{x}$, licensing occurs if $x^* \leq x < \infty$ and the outside option is infringement because $x > \widetilde{x}$. So, if $x^* > \widetilde{x}$,

$$G(x,\theta) = p(1-\lambda)t(x,\theta) - c^P - c^A.$$

So, in either situation, $G(x^*,\theta) = 0$ and we have $\frac{dx^*}{d\theta} = \frac{-t_\theta(x^*,\theta)}{t_x(x^*,\theta)} < 0$ and $\frac{d(1-F(x^*(\theta))).}{d\theta} = -f(x^*(\theta))\frac{dx^*}{d\theta} > 0$. If $x^* = \widetilde{x}$, there is a discontinuity and it occurs at an upward jump,

$$G^+(\widetilde{x},\theta) - G^-(\widetilde{x},\theta) = (1-p)(1-\lambda)t(x,\theta) > 0.$$

If $x^* = \widetilde{x}$, $G(x^*,\theta) > 0$ and the outside option is infringement for $x^* \leq x$. Because of the discontinuity, there is a range of θ such that $x^* = \widetilde{x}$, so that inside this range, $\frac{dx^*}{d\theta} = \frac{d\widetilde{x}}{d\theta}$. From $\Pi_0^A(\widetilde{x},\theta) = \pi_0^A$, it follows that $\frac{d\widetilde{x}(\theta)}{d\theta} = -\frac{t_\theta(\widetilde{x},\theta)}{t_x(\widetilde{x},\theta)} < 0$, which implies that $\frac{dx^*}{d\theta} < 0$ for $x^* = \widetilde{x}$ and $\frac{d(1-F(x^*(\theta)))}{d\theta} = -f(x^*(\theta))\frac{\partial x^*}{\partial\theta} > 0$. It can be shown that $\frac{dx^*}{d\theta} < 0$ holds at the endpoints of the range of θ where $x^* = \widetilde{x}$. An increase in θ causes the outside options to shift from those with infringement to those without technology transfer. An increase in θ cannot cause the outside options to shift from those without technology transfer to those with infringement. It follows that $\frac{dx^*}{d\theta} < 0$ and $\frac{d(1-F(x^*(\theta))).}{d\theta} > 0$. $\square$

Proof of Proposition 3. The probability of adoption is given by $\Pr(\text{adoption}) = 1 - F(\min\{\widetilde{x}(\theta), x^*(\theta)\})$, which is increasing in θ if $\widetilde{x}(\theta) \geq x^*(\theta)$ by Proposition 2. If $\widetilde{x}(\theta) < x^*(\theta)$, $\frac{d\widetilde{x}(\theta)}{d\theta} = -\frac{t_\theta(\widetilde{x},\theta)}{t_x(\widetilde{x},\theta)} < 0$. It follows that $\frac{\partial(1-F(\min\{\widetilde{x}(\theta),x^*(\theta)\}))}{\partial\theta} > 0.\square$

B Data Appendix

In the document "Matching assignees and assignors in the USPTO Patent Assignment Dataset to Compustat firms" we describe the creation of the ASSIGNEE/OR-GVKEY file which matches assignor/ee names in the USPTO Patent Assignment Dataset to Compustat GVKEYs. This file almost directly produces the dataset on patent trades. Regarding the other forms of technology transfer, we match the names of the parties in the Licensing, Cross-licensing and R&D alliances deals to assignor/ee names in the ASSIGNEE/OR-GVKEY file.

Patent trades Using the ASSIGNEE/OR-GVKEY link we retain "assignments of the assignor's interest" in the USPTO Patent Assignment Dataset in which at least one of the assignors and one of the assignees are linked to a GVKEY. If assignors/ees are linked to multiple GVKEYs we use the relevant GVKEY during the execution date of the assignment. We then create a file with all the assignee-assignor interactions keeping just one observation per pair and removing assignments in which the GVKEY of the assignee and the assignor are the same (this happens when patents are reassigned between firms belonging to the same corporate group). The final file contains 4,495 unique GVKEY interactions taking place between 1981 and 2013.

Licensing We purchased a dataset from ktMINE including, among other information, the name(s) of the licensor(s) and the licensee(s) of 12,122 licensing deals (some of which could be duplicates) extracted from SEC filings. We cleaned and harmonized licensor and licensee names obtaining 12,304 unique names. We were able to match 3,794 of these names to assignee/or names in the ASSIGNEE/OR-GVKEY file. We retained deals with at least one licensor and one licensee linked to a GVKEY finding 3,795 unique GVKEY interactions.

We also downloaded 14,270 alliances with a licensing agreement flag from Thomson Reuters Joint Venture & Strategic Alliances Database. Name cleaning and harmonization yielded 12,976 unique names. We were able to match 2,910 of these names to assignee/or names in the ASSIGNEE/OR-GVKEY file. We retained deals with at least one licensor and one licensee linked to a GVKEY finding 3,938 unique GVKEY interactions.

Overall, there are 9,833 GVKEY interactions (some interactions are both in the ktMINE and Joint Venture & Strategic Alliances Database databases).

Cross-licensing We created a list of cross-licensing deals by carrying out an exhaustive search across forms disclosed to the SEC. We downloaded all the SEC forms filed from 2000 to 2014 (both inclusive) containing the word "cross-licensing" (or related strings such as "cross licensing", "cross-license" or "cross license"). This resulted in approximately 22,500 forms (mainly 10-K and 10-Q, but also 424B3, 8-K, S-4 and other types of forms). We carefully read each one of these forms and extracted information on all the 4,375 instances in which the identity of the parties in the agreement was disclosed. Some of these refer to repeated cross-licensing deals disclosed by filers year after year. We complemented this list with a Google search over two Compustat samples that are likely to engage in cross-licensing. The first one of these samples comprises 1,482 firms with an average patent stock of more than 20 patents. The second one includes 1,213 firms with no patents but with average yearly R&D expenditures above $3 million. We searched for the name of the selected companies together with the word "cross-licensing" obtaining 599 cross-licensing deals with information on, at least, the names of the cross-licensees. We appended the SEC and Google searches together and harmonized the names of the cross-licensees finding 2,608 unique names. We were able to match 1,492 of these names to assignee/or names in the ASSIGNEE/OR-GVKEY file. We retained deals with at least two cross-licensees linked to a GVKEY finding 1,589 unique GVKEY interactions. The flow of technology transfer is bidirectional in cross-licensing agreements. This implies that the number of knowledge adopters through cross-licensing is 3,178 (1,589*2).

Additionally, we downloaded 8,434 alliances with a cross-technology transfer agreement flag from Thomson Reuters Joint Venture & Strategic Alliances Database. Name cleaning and harmonization yielded 9,397 unique names. We were able to match XXX of these names to assignee/or names in the ASSIGNEE/OR-GVKEY file. We retained deals with at least two cross-licensees linked to a GVKEY finding 2,371 unique GVKEY interactions. The flow of technology transfer is bidirectional in cross-licensing agreements. This implies that the number of knowledge adopters through cross-licensing is 4,742 (2,371*2).

Overall, there are 3,812 GVKEY interactions and 7,622 adopters (some interactions are both in the ktMINE and Joint Venture & Strategic Alliances Database databases).

R&D alliances We downloaded all the 16,160 R&D alliances available in SDC platinum in March 2016. We cleaned and harmonized the names of the firms forming the alliance obtaining 13,576 unique names. We were able to match 2,814 of these names to assignee/or names in the ASSIGNEE/OR-GVKEY file through CUSIP numbers and harmonized names. We retained R&D alliances where at least two of the firms were successfully linked to a GVKEY. This resulted in 4,486 unique GVKEY interactions. The flow of technology transfer is bidirectional in R&D alliances. This implies that the number of knowledge adopters through R&D alliances is 8,972 (4,486*2)