Working Paper

No 2024-08 – July

CEPII

The Percolation of Knowledge across Space

Pierre Cotterlaz & Arthur Guillouzouic

Highlights

- We show that the negative effect of geographical distance on knowledge flows stems from how firms gain sources of knowledge through their existing network.
- In aggregate, the distance elasticity of patent citations flows has remained constant since the 1980s, despite the rise of the internet.
- At the micro level, firms disproportionately cite existing knowledge sources, and patents cited by their sources.
- We introduce a model featuring the latter phenomenon, and generating a negative distance elasticity in aggregate.
- While the distance effect should have decreased with constant country composition, the rise of East Asian economies, associated to large distance elasticities, compensated lower frictions in other countries.

RESEARCH AND EXPERTISE ON THE WORLD ECONOMY

Abstract

This paper shows that the negative effect of geographical distance on knowledge flows stems from how firms gain sources of knowledge through their existing network. We start by documenting two stylized facts. First, in aggregate, the distance elasticity of patent citations flows has remained constant since the 1980s, despite the rise of the internet. Second, at the micro level, firms disproportionately cite existing knowledge sources, and patents cited by their sources. We introduce a framework featuring the latter phenomenon, and generating a negative distance elasticity in aggregate. The model predicts Pareto-distributed innovator sizes, and citation distances increasing with innovator size. These predictions hold well empirically. We investigate changes of the underlying parameters and geographical composition effects over the period. While the distance effect should have decreased with constant country composition, the rise of East Asian economies, associated to large distance elasticities, compensated lower frictions in other countries.

Keywords

Knowledge Diffusion,Innovation Networks, Spatial Frictions, Patent Citation.

JEL

L14, O33, R12.

Working Paper

CEPII

Centre d'études prospectives
et d'informations internationales
20, avenue de Ségur
TSA 10726
75334 Paris Cedex 07

contact@cepii.fr
www.cepii.fr – @CEPII_Paris
Press contact: presse@cepii.fr

CEPII Working Paper
Contributing to research in international economics

CEPII (Centre d'Études Prospectives et d'Informations Internationales) is a French institute dedicated to producing independent, policy-oriented economic research helpful to understand the international economic environment and challenges in the areas of trade policy, competitiveness, macroeconomics, international finance and growth.

Editorial Director:
Antoine Bouët

Visual design and production:
Laure Boivin

ISSN 2970-491X

July 2024

To subscribe to
The CEPII Newsletter:
www.cepii.fr/KeepInformed

RESEARCH AND EXPERTISE
ON THE WORLD ECONOMY

1 Introduction

While innovation and knowledge spillovers are a key engine of economic growth (Aghion and Jaravel 2015), these spillovers seem to have weakened in the recent past, and ideas seem ever harder to find (Akcigit and Ates 2021; Bloom et al. 2020). Although these trends may partly be due to firms actively trying to avoid knowledge diffusion in order to protect their competitive advantage, the main macro-level impediment to knowledge dissemination remains the mere geographical distance between innovators. Thus, it is crucial to understand the forces that underlie the imperfect dissemination of knowledge, both between innovators and across space, and how these have been affected by the tremendous progress of information search technologies achieved in the past four decades.

This paper studies the network formation underpinning the negative impact of geographical distance on knowledge flows. We first document two critical stylized facts. At the macro level, the elasticity of patent citation flows with respect to distance has remained remarkably stable since the 1980s, despite the rise of the internet. At the micro level, we observe a pattern of knowledge percolation within networks, with innovators disproportionately citing sources of their own sources. We then develop a model that bridges these observations, incorporating a network formation process among innovators, and generating an aggregate effect of distance. The model delivers two predictions: the size of innovators should be Pareto-distributed, and a systematic relation should link firms' size and the distance of their citations. Meeting these two predictions in turn generates a negative distance elasticity of knowledge flows with respect to geographical distance, our initial stylized fact. The data fit these predictions well, and changes in the predicted parameters correlate with changes in the distance elasticity. To investigate drivers of the distance effect dynamics, we decompose it along two dimensions: technologies and countries. We find that, while changes in the technological composition of innovation over the period had no impact on the overall distance elasticity, the geographical composition of innovators mattered considerably. The effect of distance should have become smaller, but the rise of China and Korea, with pronounced distance effects, eventually offset these gains. This suggests different patterns of knowledge acquisition in these countries, which we hypothesize to be partly due to language barriers.

We start by documenting that, over almost 40 years (1980–2017), the elasticity of citation flows with respect to distance has remained remarkably stable. As is standard for the study of trade flows, we adopt a gravity equation framework (Head and Mayer 2014), and find a distance elasticity of -0.21, around a fifth of the distance elasticity of trade flows. While previous studies have documented the existence of a negative distance elasticity of citation flows (Maurseth and Verspagen 2002; Peri 2005; Griffith, Lee, and Van Reenen 2011; Li 2014), these studies each focus on only one patent office and stop around year 2000. We present broader evidence, encompassing all patent offices, and extending the existing results by almost two decades. Studying the recent period is important, as it was marked by the generalization of the use of internet and the onset of search tools such as Google patents, which could have been expected to greatly limit search frictions, thus decreasing the effect of distance. In contrast, our findings document that the distance elasticity in the 2010s is no different from the one measured in the 1980s.

In the second stylized fact, we document a micro-level pattern of link formation, in which firms disproportionately form links with sources of sources. This diffusion process is reminiscent of the physics phenomenon of percolation, approaching knowledge as a fluid making its way from one firm to another along network paths. Specifically, we observe that once a link is established, innovators are more inclined to cite patents from both the source firms and those cited by these sources. To demonstrate this phenomenon, we devise a test for diffusion along network links, contrasting realized citations with a frictionless scenario where innovators would cite every relevant patent. We utilize examiner-added citations, whose only difference with applicant citations arguably is that the applicant was not aware of them, to construct a counterfactual set of citations that would arise in a frictionless world where all relevant patents would be cited. By allowing us to compare observed citations to this counterfactual, this setting identifies the effect of awareness links within the innovators' network on knowledge utilization. Our analysis reveals that firms are more likely to cite patents from their sources compared to those from outside their network, indicating the persistent value of these links. Furthermore, this effect extends beyond direct links, with patents cited by at least one of the firm's sources being more likely to be cited. Our findings withstand various robustness tests, including checks for potential strategic citation omissions or relevance differences between applicant and examiner-added references.

We present a model able to bridge the above two facts: starting from a micro network formation process, this model generates a negative effect of distance in aggregate. To do so, we adapt Chaney (2018)'s model to the context of knowledge diffusion. In this dynamic model, newborn firms have spatially clustered sources and gain new sources through their network, adapting the established idea of triadic closure in the social networks literature, *i.e.* the disproportionately high likelihood to make friends with friends of friends (Jackson and Rogers 2007). The model generates two predictions. First, the distribution of firm sizes (number of knowledge sources) should be Pareto. Second, a systematic relation links the size of firms and the distance at which they cite, reflecting the spatial distribution of their sources. Under mild conditions, these two facts can be combined to generate a negative distance elasticity, which is then a function of the shape parameter of the size distribution and of the parameter linking size and (squared) distance of citations.

Finally, we confront the theoretical predictions to the data and assess the model fit. We find that the size distribution of innovators fits a Pareto law very well, with a shape parameter (called λ) slightly above 1, making it enter the wide class of objects following a Zipf law (Gabaix 2016). The second prediction also holds: the (squared) distance at which firms cite is systematically associated to their size (with a slope denoted μ). Moreover, disaggregating our sample across countries, technologies and time periods, we find that the measured distance elasticity varies with the estimated parameters λ and μ emanating from the network formation model, accordingly to what theory predicts. We then investigate the stability of the distance elasticity, and use the equivalence between the distance elasticity, λ and μ provided by the model to assess the contribution of changes in the composition of technologies and countries over the period. We show that overall changes in the parameters λ and μ underlying the effect of distance, stemming from network search, are consistent with the observed stability of ζ. While λ decreased over the period, reflecting an increased concentration of innovation which lowered the effect of distance, μ also decreased, compensating the former effect. Moreover, we find that changes in the technological composition had no impact on the effect of distance, but that the geographical composition mattered considerably. In particular, while the effect of distance should have weakened, had the country composition remained constant, the rise of East Asian economies with large distance effects counterbalanced the trend.

An important takeaway of this paper is that small firms are important contributors to the aggregate effect of distance. Innovators start off relying on knowledge produced by sources located close to them, and get links with innovators located further away as they grow through network search, progressively escaping gravity as they grow older. This entails natural policy implications: if providing firms with all the necesary knowledge makes them closer to the innovation frontier, innovation policy should focus on exposing firms to various knowledge sources. To achive this goal, programs implying R&D collaboration may allow exchanges between small and large firms, widening small firms' horizon, while standard policy instruments such as direct subsidies or research tax credits may increase small firms' efforts but are unlikely to broaden the set of knowledge they rely on, and cluster policies may simply allow forming more links but with spatially very close firms.

Our paper contributes to several strands of the literature. First, our work contributes to a line of work studying the impact of geographical distance on knowledge flows both at the micro (Jaffe, Trajtenberg, and Henderson 1993; Thompson and Fox-Kean 2005; Thompson 2006; Murata et al. 2014) and aggregate levels (Maurseth and Verspagen 2002; Peri 2005; Griffith, Lee, and Van Reenen 2011; Li 2014). Compared to the latter, our estimates cover all patent offices and extend the series to the recent period (2000s and 2010s), a crucial addition considering the deep changes that occurred in the technologies to search and exchange ideas during this time.

Second, our work contributes to a literature studying the interplay between the effect of network proximity and the effect of distance on technological knowledge diffusion. Most of the existing literature has shown that the effect of distance is eroded when one accounts for the network structure of strong ties between inventors (Singh 2005; Kerr 2008; Agrawal, Kapur, and McHale 2008; Breschi and Lissoni 2009). These measures of social ties include co-patenting, mobility of skilled workers, or proxies such as common ethnicity. Focusing on scientists, Head, Li, and Minondo (2019) study citations between articles in mathematics, and control for social ties in an elaborate way, building connections based on past acquaintances (working in the same institution, being one's PhD supervisor, etc.). They find that controlling for ties halves the coefficient associated to distance, which is small and insignificantly different from 0 in the recent period. In contrast to the above strand of the literature, we do not restrict our attention to a particular type of links between inventors. Instead, we use awareness links as a measure of weak ties, and build these links between innovating firms rather than inventors. Our approach starts from the formation of weak links between firms, and shows how the network formation process generates an effect of distance in aggregate.

Our study also contributes to a literature providing micro foundations to aggregate geographical frictions observed on trade and knowledge flows (Buera and Oberfield 2020; Chaney 2014; 2018). By applying the framework of Chaney (2018) to a new object, we uncover important similarities between trade flows and knowledge flows. This allows us to document novel facts on innovators, such as the Pareto distribution of their sizes, or a systematic link between the size of an innovator and the distance at which is cites, providing general evidence to insights from an earlier case study (Almeida and Kogut 1997).

The remainder of the paper proceeds as follows. Section 2 describes our data source and develops the two stylized facts on which we build the analysis. Section 3 develops a model bridging our two stylized facts. Section 4 brings predictions of the model to the data and constructs counterfactuals, and section 5 concludes.

2 Data and Stylized Facts

2.1 Data

Our study relies on the data source Patstat, Spring 2022 edition. Patstat is produced by the European Patent Office and covers patent applications in almost all patent offices in the world. Given its scope, it is one of the most widely used sources on patents. In this subsection, we describe key features of the data and concepts we use throughout the paper.

Patent Citations. The standard approach in the literature to track knowledge flows has been the use of patent citations: when applying for a patent, the applicant is required to cite relevant prior art on which its invention builds. Patent citations added by applicants have been shown to reflect awareness and a potential knowledge transfer from the cited patent to the citing patent[1], albeit imperfectly.[2]

1. See surveys such as Jaffe, Trajtenberg, and Fogarty (2000), Duguet and MacGarvie (2005), or Corsino, Mariani, and Torrisi (2019).

2. Drawbacks of this measure include the fact that many patents are valueless, that citation rules vary across offices, that citations can be handled by lawyers rather than inventors, include some strategic considerations (Lampe 2012; Cotropia, Lemley, and Sampat 2013; Corsino, Mariani, and Torrisi 2019), that the use of citations could have changed over time (Kuhn, Younge, and Marco 2020) or that different industries use patents applications with different purposes (Corsino, Mariani, and Torrisi 2019).

Patent citations are of two types: applicant citations (henceforth AA) and examiner citations (henceforth EA). These are added according to the following procedure. At the time of the application, patent assignees are asked to cite the relevant prior art,[3] which helps judge the patentability of the invention, and notably its novelty relative to the existing technological background. After that, an office examiner assesses novelty of each of the claims that the patent contains, and looks for relevant prior art with the variety of tools at her disposal, adding the references which are relevant for patentability.[4]

Therefore, an important point to keep in mind is that only applicant-added citations are likely to reflect a transfer of knowledge and awareness from the inventors. In constrast, because they are added by a third party after the invention process, examiner-added citations seem very unlikely to change the set of patents an applicant is aware of. In other words, examiner citations will not contribute to the definition of knowledge sources given below. An important fact is however that, although the process is sequential, the reference list of the examiner is established independently of the applicant's list.[5] An illustration of this is the considerable overlap of citation lists: indeed, 20% of the 47 million citations made at the USPTO since 2000 are made both by the examiner and the applicant. Another very common phenomenon is self-citation, *i.e.* a citation pointing to a previous patent of the assignee applying for the patent, which we naturally exclude throughout the paper as they are unlikely to reflect knowledge transfers.[6]

Patent Applicants. Since our focus is on firms, we consider patent applications and citations at the level of the patent assignee, i.e. the physical or legal person owning the property rights over the invention. We provide more details about the content of the data in Section A of the Appendix: in particular, subsection A.2 further describes how we identify assignees and their country of origin.

2.2 Stylized Fact #1: the Persistent Effect of Distance on Knowledge Flows

While the mere fact that distance negatively affects patent citation flows is well-established (Maurseth and Verspagen 2002; Peri 2005; Griffith, Lee, and Van Reenen 2011; Li 2014), exist-

3. At the USPTO, 35 U.S. Code § 301 a) requires that "Any person at any time may cite to the Office in writing (...) prior art consisting of patents or printed publications which that person believes to have a bearing on the patentability of any claim of a particular patent." The exact nature of this requirement varies slightly across offices: for instance, applicants at the USPTO have the obligation (called "duty of candor") to do so for the patent to be enforceable once granted, while the requirement is softer yet sufficient to preserve incentives to cite at the EPO (Akers 2000).

4. "Upon creation of a European search report [...], a pre-search algorithm generating a list of documents to be inspected by the examiner is triggered.[...] The examiner should start the search process by formulating a search strategy, i.e. a plan consisting of a series of search statements expressing the subject of the search, resulting in sections of the documentation to be consulted for the search." (EPO 2016)

5. Cotropia, Lemley, and Sampat (2013) indeed show that examiners completely ignore applicants' reference lists while building their own, making both lists completely independent.

6. We consider an outward citation to be a "self-citation" as soon as the cited and the citing patent have at least one common applicant or inventor.

ing studies typically stop at the end of the 1990s, before the time when one might expect the absolute effect of distance to fall due to the rise of the internet and much improved search technologies. We therefore test for the existence of spatial frictions in the diffusion of knowledge, from the 1980s to the end of the years 2010s. We do so in a standard way, studying the sensitivity of the flows of outward patent citations to distance in a simple gravity framework (Head and Mayer 2014). Geographical distance between countries comes from the CEPII GeoDist dataset.[7]

Data is aggregated at the $tijk$ level, where t denotes the year of the patent application, i denotes the country of the citing applicant, j the country of the cited applicant, and k the wide technological class (IPC at the 1-digit level).

To obtain the yearly distance elasticities, we estimate:

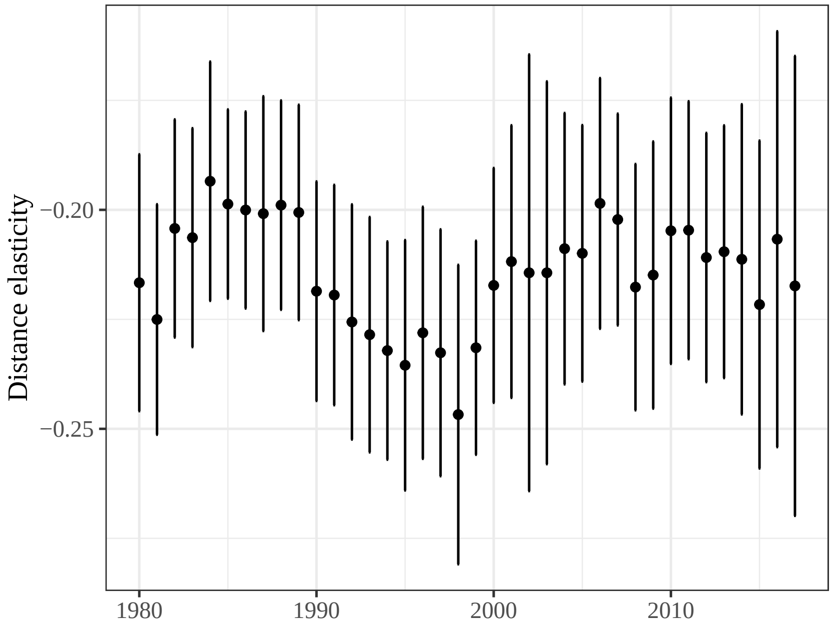
$$Y_{tijk} = \exp\left(\sum_{s=1980}^{2017} \zeta_s \cdot \text{dist}_{ij} \cdot \mathbb{1}(t=s) + \Omega_{tik} + \Theta_{tjk} + \varepsilon_{tijk}\right) \tag{1}$$

where Ω_{tik} is a citing country $\times$ technology $\times$ year fixed-effect and Θ_{tjk} is a cited country $\times$ technology $\times$ year fixed-effect, jointly capturing any time-varying characteristics at the level of the country (respectively citing and cited) and technology, such as the intensity and quality of innovation in that unit. ε_{tijk} is a residual.

Figure 1 shows the evolution over the period 1980–2017 of our yearly ζ estimate. The stability of the coefficient over the period is very striking: the elasticity of citations flows with respect to distance is close to -0.21 over the whole period, with only the 1990s showing a slight decrease to around -0.24, implying that the effect of distance temporarily became stronger in this decade. This is confirmed by Table 1, showing differences between estimates across decades. The column corresponding to decade 1990 shows small yet significant differences with the rest of the period, implying an elasticity around 10% larger in absolute terms in this decade. Importantly, all other differences are very close to 0 and insignificant statistically, showing that the effect in the years 2000s and 2010s, a period in which efficient patent search tools were widely available, is the same as in the 1980s, before the adoption of the internet.

7. See Mayer and Zignago 2011, http://www.cepii.fr/cepii/fr/bdd_modele/presentation.asp?id=6.

Figure 1: Evolution of the distance elasticity ζ over time



NOTES: We estimate yearly distance elasticities of patent citation flows from equation (1) using a PPML estimator. The sample pools all patent offices. Intra-national citations and firm self-citations are excluded. The distance between countries corresponds to the distance between the largest city of each country, obtained from the CEPII Geodist dataset. Bars display the 95% confidence interval for each estimate. Standard-errors are clustered at the “year $\times$ country $\times$ technology" level (both for the origin and the destination country).

Table 1: Difference between decades for the distance elasticity ζ

Ref. decade	**Decade** 1980	1990	2000	2010
1980	–	-0.0303*** (0.0067)	-0.0084 (0.0069)	-0.0087 (0.0079)
1990	0.0303*** (0.0067)	–	0.0219*** (0.0073)	0.0216* (0.0083)
2000	0.0084 (0.0069)	-0.0219*** (0.0073)	–	-0.0003 (0.0084)
2010	0.0087 (0.0079)	-0.0216* (0.0083)	0.0003 (0.0084)	–

NOTES: Distance elasticities by decade are estimated from: $Y_{tijk} = \exp\left(\sum_{\text{decade}=1980}^{2010} \zeta_{\text{decade}} \cdot \text{dist}_{ij} \cdot \mathbb{1}(t \in \text{decade}) + \Omega_{tik} + \Theta_{tjk} + \varepsilon_{tijk}\right)$, using a PPML estimator, similarly to equation (1). Values are the difference between the distance elasticity estimate for the decade in columns and the distance elasticity for the decade in rows.

2.3 Stylized Fact #2: Network Formation through Existing Knowledge Sources

The existence of a negative distance elasticity, even after accounting for all characteristics of origins and destinations, shows that there exist frictions which prevent innovators from knowing about (and therefore citing) all relevant references. In this subsection, we seek to understand why we depart from a frictionless context in knowledge diffusion by studying innovators' citation behavior. In particular, we want to characterize the influence of existing knowledge sources, as evidenced by past citations, on new link formation. We test this along two dimensions. First, we measure whether innovators have a disproportionate tendency to cite new patents (i.e. patents they have never cited), belonging to existing sources (i.e. firms they have already cited). Second, we measure whether innovators are prone to citing patents belonging to sources of sources, that is, patents already cited by innovators they know.

The first hypothesis we test amounts to testing whether the links we define between a citing firm and a knowledge source are persistent, in the sense that knowledge is likely to flow again along a link where it has flowed once. The second hypothesis we test is whether knowledge diffuses in the network through the creation of new links, and in particular two steps away (at degree two in the network) from the citing firm.

2.3.1 Empirical strategy.

Counterfactual list of citations. Our goal is to compare the realized citation behavior of patent applicants to a frictionless counterfactual in which applicants would cite not only the useful patents they are aware of, but all patents which are relevant to the invention. Therefore, in order to measure deviations from randomness for each citing patent, we need a set of patents which are relevant to the patented invention, but which the applicant was not necessarily aware of. Citations added by office examiners provide an ideal list of patents above a certain level of relevance to the patented invention. As presented in subsection 2.1, these patents are added by field experts, and are on average a wider list of references than those added by the applicant but with a large degree of overlap. We consider the union of both lists as the full set of citable patents over a given level of relevance, and study if realized applicant citations depart from this counterfactual situation in systematic ways. In practice, we compare the set of applicant-added citations AA to the set of examiner citations which were not made by the applicant, EA \ AA.

Identifying assumptions. Our strategy relies on the fact that applicant and examiner citations are comparable in all respects, except for the fact that the applicant knew about the cited patent. An assessment of their observable characteristics, plotted in Figure A4 in the Appendix, supports the idea that they are very comparable: AA and EA citations are almost exactly similar in terms of technological and geographical distance, and EA citations only seem slightly younger and of lower quality, which we can easily control for. Regarding unobservable characteristics, the identifying assumptions we make are the following. We think of two key unobservable features: relevance to the citing patent, and awareness of the person making the citation. In a nutshell, the validity of our comparisons rests on the assumption that the only unobserved characteristic along which patents in the two groups (applicant and examiner-added) differ is whether the applicant was aware of them or not. This assumption has two corollaries.

First, there should be no systematic differences of relevance between patents cited by examiners and by applicants. While this assumption cannot be tested directly, several pieces of evidence support it. As shown in Appendix A.4, matching USPTO applications to the PAIR dataset allows us to study examiners' behavior. Overall, our description suggests that examiners tend to be experts in their field, carry an extensive and independent search on relevant existing patents, and be little influenced by past searches they may have done.[8] Moreover, as mentioned above, examiner citations overlap applicant citations to a large degree. Beyond the fact that this feature reveals great similarity in relevance, we can use it to conduct checks on the alternative sample consisting only of examiner citations, within which some citations are also made by applicants.

The second corrolary is that missing citations to relevant patents by applicants can only stem from absence of awareness. In other words, applicants always have an incentive to cite any relevant patent they know, because it strengthens their application and that the examiner would find other relevant patents in any case. This is of course a simplification, and neglects the possibility for applicants to strategically withhold some citations. Lampe (2012) notably shows evidence that strategic withholding is frequent. While this is likely to bias our network effect estimates downward, we follow Lampe (2012)'s definition of strategic citations and design a test to handle them.

Construction of network links. To implement the above-mentioned tests, we start by defining links in the network at a given point in time. We start from origin firms (firms applying for a patent) and define as their source firms all the assignees of patents truly cited (i.e. applicant citations to these firms). We exclude citations to industry leaders (defined as firms in the top 1% of patent applicants) in the formation of links. The logic behind this choice is that, beyond the computation cost they imply, patents applied for by very large company are widely visible and can be known by all without search frictions. We however provide evidence that changes in the definition of industry leaders (ranging from the top 10% to the top 0.01%) do not affect our conclusions.

We then build citation links of distance 2 in the network of any given applicant A, meaning that such patents are two steps away from applicant A: they have been cited by an applicant B, which belong to applicant A's sources. Distance 2 links therefore define the sources of sources.

8. We match approximately 5 million USPTO applications with examiner information to USPTO's Public PAIR data. On average, examiners seem to be specialized in a field, display little persistence in their behavior, and do not lose accuracy when they do cite a patent several times. Moreover, as shown by Lei and Wright (2017), the fact that a thorough search has been conducted is true regardless of the value of the patent.

Note that these links are directed: the fact that A cites B implies a knowledge transfer from B to A, but has no implications for transfers from A to B.

Estimation sample. Once links between origin firms and their sources are formed, at a given point in time, our tests imply studying the subsequent citation behavior of firms. For all years between 2000 and 2015, we define source firms in a given year as all the firms cited in that year. Sources of sources are then all the firms which have previously been cited by sources. Building distance 2 links is computationally demanding: to alleviate computations while keeping high statistical power, we randomly select for each initialization year a third of all firms which both patent in that year and in a subsequent year.

For each year initializing network links, our sample is made of all citations from subsequent patents applied for at the USPTO by origin firms. These citations may be applicant added or not (which defines our dependent variable), and may be citing patents belonging to source firms (which defines our first variable of interest), or to sources of sources (defining our second variable of interest). Importantly, we always control extensively for whether cited patents or applicants have already been cited in years prior to the initialization year, implying that only new links from the initialization year contribute to the estimation of our coefficients of interest.

To obtain our final sample, we then stack each subsample associated to a cohort between 2000 and 2015 when links are initialized. This implies that citations may appear several times in the sample, although with different values taken by the network variables, because links are cohort-specific. A drawback of this approach is that patents applied for late in the period have a higher probability of appearing several times in the sample. We conduct robustness imposing a window of three or five years after the initialization year. Our final sample contains close to 65 millions of citations made in 2.2 millions of patent applications by more than 460 thousand firms. Figure B1 in the Appendix provides a graphical depiction of our sample and regressors construction.

Specification. In our preferred model, we estimate, through ordinary least squares, the following specification:

$$Y_{odc} = \beta_1 \cdot L^{(1)}_{odc} + \beta_2 \cdot L^{(2)}_{odc} + \boldsymbol{\gamma} \cdot \mathbf{X}_{od} + \boldsymbol{\zeta} \cdot \mathbf{P}_{odc} + \nu_{oc} + \eta_{f(d)} + \varepsilon_{odc} \tag{2}$$

where o denotes a citing (origin) patent, d a cited (destination) patent, c the year of contact initialization (cohort). Y_{odc} is a dummy variable indicating whether patent o cites (i.e. through an AA citation) patent d, $L^{(1)}_{odc}$ indicates a link of distance 1 (belonging to a knowledge source) from patent o to patent d as of initialization year c, $L^{(2)}_{odc}$ indicates a link of distance 2 (having been cited by a source) from patent o to patent d as of initialization year c, $\mathbf{X_{od}}$ is a set of observable characteristics at the citation level, $\mathbf{P}_{odc}$ is a set of control variables indicating past citation links between o and d (whether at the firm, patent family or patent level), ν_{oc} is an origin patent $\times$ initialization year fixed-effect, $\eta_{f(d)}$ is a destination firm fixed-effect, and ε_{odc} is a residual.

A considerable advantage of using OLS to estimate our model is that it allows for the introduction of a rich set of fixed-effects in the regression, while remaining easy to estimate on very large samples such as ours. In our context, this implies being able to add destination firm fixed-effects in our specification, therefore capturing all fixed characteristics from the destination assignee in terms of country, age, quality and technological composition of the patent portfolio, to focus only on the pairwise variation between the origin and the destination firm. More traditional discrete choice models, such as the conditional logit, are typically very computationally demanding on very large samples, but have the advantage of having a direct mapping to theory. We favor the former to the latter in our baseline exposition of the results, but provide evidence that conditional logit estimates provide similar results (although without estimating the $\eta_{f(d)}$).

2.3.2 Results

Table 2 presents the main results of our analysis, varying the specification used. In all specifications, regressions include a cohort-specific citing patent fixed-effect, and a set of control variables capturing all the citation behavior of applicants prior to the cohort year. Our preferred specification, corresponding to equation (2), is presented in column (4), and also includes cited firm fixed-effects, capturing any relevant dimensions of quality, visibility, etc., of the firm cited by patents, as well as pairwise (citing-cited patent) controls, including the difference in quality, age, technological and geographical distance.

Two conclusions emerge from this table. First, the fact of being a patent belonging to a source makes the occurrence of a (AA) citation 2 pp more likely, corresponding to a 3.2% increase. This effect is remarkably stable in magnitude across specifications, and implies that, once a link is established through an initial citation, knowledge is more likely to flow again along that link through citations toward "new" patents of the source firm. This finding confirms that links built through citations are meaningful and tend to be persistent. Second, and most importantly, our results imply that having been cited by a source, i.e. being a source of source, also increase the likelihood that a citation occurs by 1.4 pp, which is a 2.2% increase. This implies that the network of sources tends to expand step by step, with patent applicants being more likely to learn, all other things kept equal, about knowledge developped by a firm which is a source of one of their own sources. This phenomenon, called triadic closure in the economics of networks literature, has deep implications on knowledge diffusion in aggregate, which we develop in the next section.

2.3.3 Robustness

As mentioned in section 2.3.1, we submit our result to a wide range of robustness checks. In particular, for each of the two threats to identification explained above, we design a specific

Table 2: Estimates of influence of existing links on citations

Dep. var.: Patent cited by the applicant				
				Baseline
	(1)	(2)	(3)	(4)
Source	0.022***	0.021***	0.021***	0.020***
	(0.0002)	(0.0002)	(0.0002)	(0.0002)
Cited by Source	0.025***	0.018***	0.021***	0.014***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Mean of the dep. variable	0.626	0.626	0.626	0.626
Number of citing firms	461.4k	461.4k	461.4k	461.4k
Number of citing patents	2.2M	2.2M	2.2M	2.2M
Number of observations	64.8M	59.5M	64.6M	59.4M
Citing patent × cohort FE	✓	✓	✓	✓
Cited firm FE	–	–	✓	✓
Pairwise controls	–	✓	–	✓
Past citations controls	✓	✓	✓	✓

NOTES: This table reports the coefficients corresponding to the specification described in Fact #2 (section 2.3). It presents regression results obtained from estimating specification (2) through ordinary least squares: the coefficient labelled "Source" corresponds to β_1, the coefficient "Cited by Source" to β_2. Standard-errors are clustered at the "citing patent × cohort" level. The estimation sample contains all patent citations from a randomly selected third of patent applicants in each given year between 2000 and 2015. The dependent variable is a dummy variable indicating if the citation was added by the applicant or not.

check. First, it may be that applicant and examiner added citations systematically differ in their relevance to the patented invention, such that our strategy does not only identify differences in awareness of existing knowledge. To handle this, we take advantage of the wide overlap between applicant and examiner citations lists. In practise, we conduct the very same tests as in the baseline, but comparing only the overlapping set to the rest of examiner-citations ($AA \cap EA$ to $EA \setminus AA$). This check ensures that the effects we measure are not driven by applicants adding irrelevant citations due to low incentives.

Second, patent applicants may withhold citations to some patents for strategic reasons, invalidating our assumption that the list of applicant citations reflects the full set of knowledge this applicant has. While such phenomenon should bias our estimates downward, we follow Lampe (2012) in spotting such citations. We define them through the fact that applicants have cited a patent in the past, showing that the applicant knew about it, but do not cite it in a further applicantion while the examiner cites it. In such case, this patent meets both the awareness and the relevance conditions that should perfectly predict a citation, yet it is not cited by the applicant. To handle this, we reclassify all patents meeting this criterion (having been cited by an applicant in the past and being cited by the examiner only later on) as patents cited by the applicant. We denote this the "patent definition" of strategic citations. We also go one step further, and tag as strategic any citation made by the examiner but not by the applicant toward a firm which had been cited in the past, and denote this the "firm definition" of strategic citations. We then conduct the same regression as before, having requalified as applicant citations the ones that were strategically withheld. While this is unlikely to reflect the whole set of strategic citations, the sensitivity of our coefficients provides hints on how serious the issue might be.

Table 3 displays the results of these two main checks. Column (2) shows our coefficients of interest when the sample is restricted to examiner citations, such that the dependent variable takes value 1 only for overlapping applicant and examiner citations. Both coefficients are approximately halved in absolute terms, but imply large effects in relative terms since the average of the dependent variable is (mechanically) much lower. Overall, we find a +0.8 pp effect associated to the fact of belonging to a source firm, corresponding to a 14% increase, and a +0.7 pp increase associated to the fact of having been cited by a source, corresponding to a 12.3% increase. This shows that we can reject the hypothesis that applicant citations tend to be of lower relevance to the patented invention than examiner citations. Columns (3) and (4) display the results obtained when we requalify citations thought to have been strategically omitted, either at the patent level (col. 3) or at the firm level (col. 4). Compared to our baseline estimate, the effect of belonging to a source firm appears to be slightly larger. The effect of having been cited by a source firm is either slightly higher or slightly lower depending on the way we implement the correction for strategic citations.

Additionally, we also conduct several robustness checks, the results of which are presented in Table B2 in the Appendix. These include the fact of changing the percentile of the size distribution above which we do not consider firms as potential sources, imposing a time window after the initialization year during which citing patents are included in the sample so as to equalize the size of cohorts in the estimation sample, and including cited patent fixed-effects instead of cited firm fixed-effects. We find that our results hold in all these contexts. The coefficient associated to belonging to a source is remarkably stable across regressions. The coefficient associated to being cited by a source remains sizeable and very significant in all specifications, only becoming a bit smaller once cited patent fixed-effects are introduced, which is expected considering how demanding this specification is.

Table 3: Main robustness checks on influence of existing links on citations

Dep. var.: Patent cited by the applicant				
	Baseline	Examiner overlap	Strategic Pat. lev.	Strategic Firm lev.
	(1)	(2)	(3)	(4)
Source	0.020***	0.008***	0.028***	0.030***
	(0.0002)	(0.0003)	(0.0003)	(0.0002)
Cited by Source	0.014***	0.007***	0.011***	0.018***
	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Mean of the dep. variable	0.626	0.057	0.652	0.665
Number of citing firms	461.4k	447k	461.4k	461.4k
Number of citing patents	2.2M	2.1M	2.2M	2.2M
Number of observations	59.4M	23.1M	59.4M	59.4M
Citing patent × cohort FE	✓	✓	✓	✓
Cited firm FE	✓	✓	✓	✓
Pairwise controls	✓	✓	✓	✓
Past citations controls	✓	✓	✓	✓

NOTES: This table reports the coefficients corresponding to the specification described in Fact #2 (section 2.3), conducting several robustness checks. It presents regression results obtained from estimating specification (2) through ordinary least squares: the coefficient labelled "Source" corresponds to β_1, the coefficient "Cited by Source" to β_2. Standard-errors are clustered at the "citing patent × cohort" level. The estimation sample contains all patent citations from a randomly selected third of patent applicants in each given year between 2000 and 2015. The dependent variable is a dummy variable indicating if the citation was added by the applicant or not. Column (1) is our baseline estimate, column (2) provides a similar estimate on a modified sample keeping only patents that were cited by examiners, columns (3) and (4) reclassify citations based on potentially strategic applicant behavior.

3 Theory: Network Origins of the Aggregate Distance Elasticity

To interpret the aggregate consequences of the micro-level empirical findings documented through Fact #2, a model featuring network formation along firms' life-cycle is warranted. This section develops a dynamic model able to bridge our finding that knowledge percolates through a network of innovators shown through Fact #2, with the fact that distance hinders aggregate knowledge flows shown as Fact #1.

To do so, we adapt a model developed in Chaney (2018) to the context of knowledge, based on the empirical findings shown in section 2.3. The mechanics are as follow: agents obtain knowledge through their sources, and start off with initial sources distributed close to them. Firms then gain some new sources as time passes, which are either the sources of their own sources (network search), or agents located close to their sources (spatial search).[9]

9. Note that the idea that young and small firms initially start with localized sources (as assumed in the model developed below) has received some empirical support: Almeida and Kogut (1997) looked at innovators in the semi-conductor industry in the US, and found that small firms were more prone to cite patents developed closer to

Model. We extend the model featured in Chaney (2018) to introduce some spatial search in addition to the network search, consistently with a vast litterature showing that knowledge spillovers have a strong spatial component, potentially implying that pure spatial forces are at play (Jaffe, Trajtenberg, and Henderson 1993; Alcácer and Gittelman 2006; Thompson 2006; Murata et al. 2014). The model is the following. Time is continuous, and infinitely-lived firms are born with a growth rate γ. They are spread uniformly in space, which is infinite and one-dimensional ($\mathbb{R}$), so that coordinates of any location are a scalar x. When they are born, firms are endowed with a set of sources of mass K_0, born at the same time, and distributed around them according to the distribution $k_0(x)$, which is assumed to be symmetric and to admit a finite second-moment. Each source provides a firm with one unit of knowledge. The set of sources of a firm of age a evolves in three ways:

- Gain via network search: a firm's existing source may reveal one of its own sources through a random Poisson shock of parameter β. This revealed source joins the set of the firm's sources. A technical constraint requires that firms can only gain sources with firms of their cohort. This corresponds to the coefficient associated to the variable **Cited by Source** in Table 2 showing that innovators do form links toward sources of sources.
- Gain via spatial search: we also allow for firms to directly find new sources, through a random Poisson shock of parameter ρ, in each location where they already have sources. This means that, going from age a to age $a + \mathrm{d}a$, the firm picks some new sources with the exact same spatial distribution as the sources it already has.
- Loss of a source, also through a Poisson shock of parameter δ.

We further assume that $\gamma > \beta - \delta > 0$ and $\gamma > \rho - \delta > 0$, which balances the relative size of small and large firms and ensures that firms' size grows as time passes. Based on these three channels, the evolution of k_a, the mass of sources at point x of an aged a firm writes:

$$\frac{\partial k_a(x)}{\partial a} = \underbrace{\rho k_a(x)}_{\text{spatial search}} + \underbrace{\beta \int_{\mathbb{R}} \frac{k_a(x-y)}{K_a} k_a(y)\mathrm{d}y}_{\text{network search}} - \underbrace{\delta k_a(x)}_{\text{source loss}} \tag{3}$$

At the same time, the evolution of the overall number of sources of a firm of age a, K_a, which is the integral of k_a, follows the simple ODE:

$$\frac{\partial K_a}{\partial a} = (\rho + \beta - \delta) K_a \tag{4}$$

with initial value K_0.

Proposition. *When the distribution of the stock and mass of sources is described by equations (3) and (4):*

them than big firms were.

- *The distribution of innovator sizes is Pareto, with a shape parameter* $\lambda = \frac{\gamma}{\rho+\beta-\delta}$*;*
- *The average squared distance at which firms cite is a power function of their number of sources, with power parameter* $\mu = \frac{\beta}{\rho+\beta-\delta}$.

Proof. See Appendix D. □

These predictions are very intuitive. In a nutshell, this model describes an environment in which firms will gradually be less and less affected by distance as they grow old: their average source is further and further away. In aggregate however, because new firms are born every period with a constant growth rate and that increases in size are generated by random shocks, this model will imply a Pareto size distribution, meaning that small firms are considerably more numerous than large ones. Moreover, because new sources are further away than old ones, the distance from sources will be an increasing function of size.

Comparative Statics. Partial derivatives of the parameters of interest with respect to ρ are as follow:

$$\frac{\partial\lambda}{\partial\rho} = \frac{-\gamma}{(\rho+\beta-\delta)^2} < 0$$
$$\frac{\partial\mu}{\partial\rho} = \frac{-\beta}{(\rho+\beta-\delta)^2} < 0$$

This means that, when spatial search increases, this generates a decrease in λ, i.e. an increase in the proportion of large firms relative to small ones. It also generates a decrease in μ, implying that the difference between the distance at which big and small firms cite drops. In other words, adding this force to the baseline Chaney (2018) model predicts a lower value for λ than if network search was the only way to gain new sources, as well as a reduced relation between firm size and distance of citations.

Similarly, partial derivatives of the parameters of interest with respect to β are:

$$\frac{\partial\lambda}{\partial\beta} = \frac{-\gamma}{(\rho+\beta-\delta)^2} < 0$$
$$\frac{\partial\mu}{\partial\beta} = \frac{\rho-\delta}{(\rho+\beta-\delta)^2} > 0$$

Thus, the effect on the distribution of firms sizes of an increase in network search is exactly equivalent to the magnitude of the effect of an increase in spatial search: it makes the tail of the size distribution thicker, by increasing the rate at which firms get new sources while leaving unchanged the entry rate of newborn firms. The sign of the effect of a change in β on μ is, however, opposite to that of a change on spatial search. Indeed, increasing the intensity of network search makes firms gain new sources further away from them compared to their set of existing sources when they grow, implying that the link between their size and the distance of their sources increases.

Gravity equation Following the proof in Chaney (2018), citation flows will exhibit a negative distance elasticity as long as the following two main conditions hold:

- **Condition 1:** Innovator sizes follow a Pareto distribution of shape parameter λ with $\lambda > 1$.
- **Condition 2:** An increasing power function with power parameter μ links the average squared distance of a firm's citations to its size.

These two conditions are exactly the predictions of the above network formation model, therefore connecting directly our stylized facts #1 and #2. Under these two sufficient conditions,[10] knowledge flows are negatively related to distance. Note also that this result is provided by normalizing $\exp(\iota)$, the average squared citation distance of the smallest firms, to 1, for the sake of simplicity.

Under these conditions, as stated above, small innovative firms are considerably more numerous than large firms, and there is a systematic relation between a firm's size and the distance of the citations it makes. In other words, if large firms cite on average further away than small firms, then citations at long distances mostly come from large firms (firms applying for many patents). Therefore, irrespective of the convergence towards an asymptotically constant elasticity of citation flows to distance, the model predicts that distance will have a smaller negative impact on patent citations (smaller ζ) if the share of large firms relative to small ones increases (smaller λ), or if the relation between size and the distance at which firms cite becomes steeper (larger μ). As in Chaney (2018), meeting these two conditions, along with an additional condition imposing that $\lambda < 1 + \mu$, implies that the elasticity of aggregate knowledge flows with respect to geographical distance is asymptotically constant and equal to $-\zeta$, with $\zeta = 1 + \frac{2(\lambda-1)}{\mu}$.

In this framework, several factors influence the aggregate effect of distance captured by ζ. In particular, an increase in network search decreases ζ in two ways: on one hand, it increases the relative share of large firms (by decreasing λ), which cite further away than small firms, while on the other hand, it makes distance of citations more dependent on size (by increasing μ), the combination of both effects in turn decreasing ζ. Moreover, while $\exp(\iota)$, the average squared distance at which the smallest firms cite, is neutralized in the above formula by a simplifying normalization, it is an intuitive margin to change ζ: because the smallest firms are very numerous, a decrease in the distance at which they cite should result in a decrease in ζ.

4 Estimation

The network formation model presented in the previous section provides sufficient conditions for a negative elasticity of knowledge flows with respect to distance to emerge. These predic-

10. Additional conditions detailed in Chaney (2018) to obtain an asymptotically constant distance elasticity are either that $\lambda < 1+\mu$ or that the PDF of citation distances of the smallest possible firm admits a finite $\left(1 + 2\frac{\lambda-1}{\mu}\right)$-th moment. More precisely, for distances going to infinity, these ensure that ζ tends towards $\left(1 + 2\frac{\lambda-1}{\mu}\right)$.

tions can be directly tested in the data. In this section, we use the number of firms cited as a proxy for size, and measure distance between applicants' countries. We find that theoretical predictions hold well empirically, which gives credence to the idea that the network formation mechanism that we described underlies the spatial decay of knowledge flows. We show that changes in the predicted value of ζ given by λ and μ within country have strong predictive power on changes in the actual distance elasticity parameter ζ. We then explore counterfactuals, seeking to understand what drove the stagnation of ζ over the period, and find that the country composition, and in particular the rise of innovation from Asian economies, is key to explain this puzzle.

4.1 Estimation of Aggregate Predictions

4.1.1 Distribution of innovator sizes (λ)

The network formation model predicts that the distribution of innovator sizes, understood as the number of sources, will be Pareto, i.e. that $F(K) = 1 - \left(\frac{K}{K_0}\right)^{\lambda}$, with $\lambda = \frac{\gamma}{\rho+\beta-\delta}$. We therefore check that a Pareto distribution fits our data well, and we estimate the shape parameter of this distribution, using the method introduced by Axtell (2001). We rank innovators by increasing order of size, where size is the number of firms cited[11] and distribute them in 20 bins of equal log width. We compute the average size of firms in each bin, denoted K_b, and the fraction of firms of size larger than K_b, denoted $1 - F(K_b)$. λ is estimated from:

$$\log[1 - F(K_b)] = a - \lambda \log(K_b) + \varepsilon_b \tag{5}$$

The absolute value of the slope of the regression line shown in Figure 2a, 1.113, corresponds to our baseline estimate of λ. The Pareto distribution fits very well our innovator size data (considering the R-squared of 98%). Since our estimated λ is very close to 1, this makes the measured distribution close to entering the specific case of a Zipf law, a Pareto distribution with shape parameter equal to 1. In the model, this implies that the net growth of the mass of sources should equate the growth rate of the firm population.

The economic literature has uncovered a wide class of objects following a power law (summarized in Gabaix 2016), which are as diverse as city sizes, income distribution, the number of trades per day, or closer to our object of study the productivity of innovations (Ghiglino 2012). We add the size distribution of patenting firms to this class. From the empirical standpoint, the distribution of firm sizes in general has been shown to follow a Zipf law by Axtell (2001). Moreover, while the assumption that productivities are Pareto distributed is common in the trade literature, Nigai (2017) has shown that the left-hand side of the distribution of productivities is closer to log-normal while the right-hand side fits the Pareto distribution better. The slight curvature of the left-hand side part of the graph suggests this might be the case as well

11. The size measure used here is the closest to the model, but all results carry through if using a more common measure of size in the growth literature, which is the number of patent applications.

for innovator sizes. However, since innovative firms are expected to have high productivity, the distribution of innovative firms sizes is likely left-truncating the productivity distribution, making the issue less salient. From the theoretical standpoint, random growth in size typically generates log-normal distributions (Gibrat 1931), while it is common to generate power laws from scale-free network formation processes (*e.g.* from the model of Albert and Barabási 2002), which also features growth in the number of nodes, and link formation through preferential attachment (which takes the form of network search in the model we use, growth alone being insufficient to generate a scale-free network).

4.1.2 Link between innovator size and distance of citations (μ)

The network model generates a second, more specific prediction. It predicts that larger firms are able to access knowledge generated further away than what smaller firms have access to. More precisely, there is a positive constant elasticity of the average squared distance at which firms cite with respect to their size. To test whether this holds in our data, we rank firms in increasing order of size,[12] and construct 20 bins of equal log width. We compute the average size of firms in each bin K_b and the average squared distance,[13] denoted Δ_b, at which firms in bin b cite. μ is estimated from:

$$\log[\Delta_b] = \iota + \mu \log(K_b) + \varepsilon_b \tag{6}$$

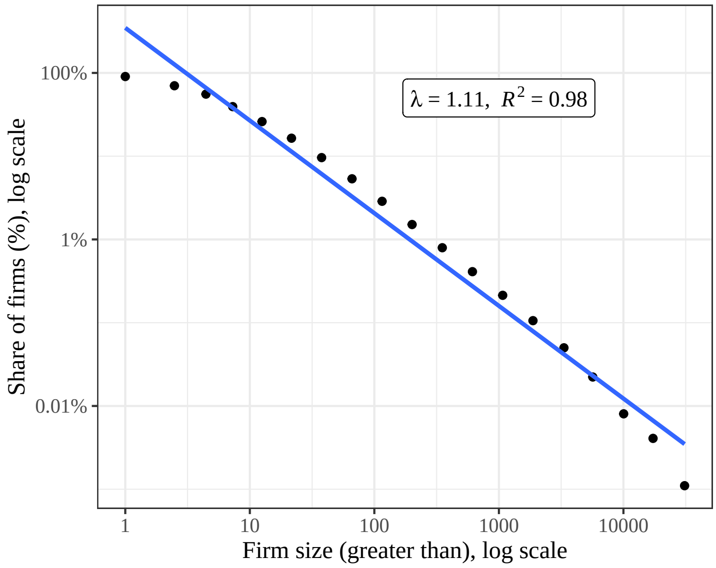
Figure 2b shows that the relationship between the average squared distance at which a firm cites and its size is well described by an increasing power function (*i.e.* increasing linear in logs). To the best of our knowledge, this systematic relationship between an innovator size and its ability to access more distant ideas is a novel finding in the analysis of patent citations. ι is the intercept of the underlying regression, and captures the (log) average squared distance at which the smallest firms are able to cite.

Taken together, these two findings are consistent with the model: older firms are larger and have links with more distant firms. Interestingly, the economy described here shares similar features with ones emanating from the Schumpeterian growth theory (e.g. Aghion, Akcigit, and Howitt 2015): the size distribution of firms, where size is assimilated to the number of their innovations, is highly skewed, and larger firms are older. Predictions relative to the distance of citations are however original and appear quite distinctive of models incorporating network formation.

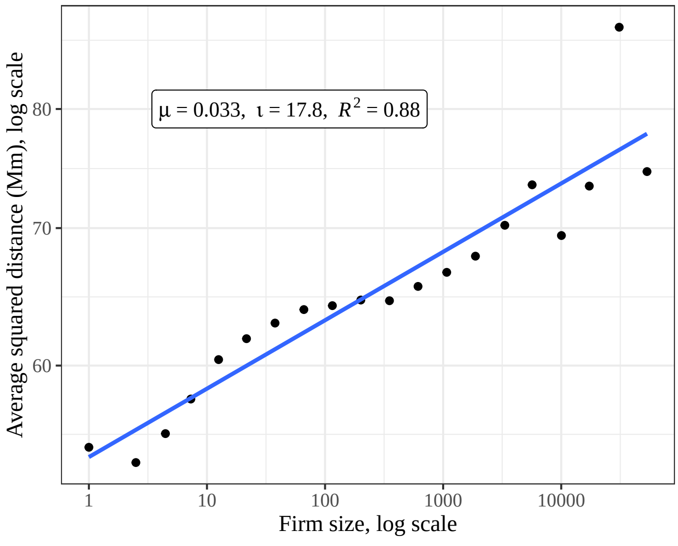
12. Size is again defined as the number of firms cited over the period 1980–2017.
13. In our baseline estimations, the distance of a citation is defined as the distance between the largest city of the country of each applicant, and intranational citations are excluded, but we show that our results hold for alternative geographic choices.

Figure 2: Estimation of λ, μ and ι.

(a) Shape parameter of the Pareto distribution of innovator sizes (λ)



(b) Link between innovator size and distance of citations (μ and ι)



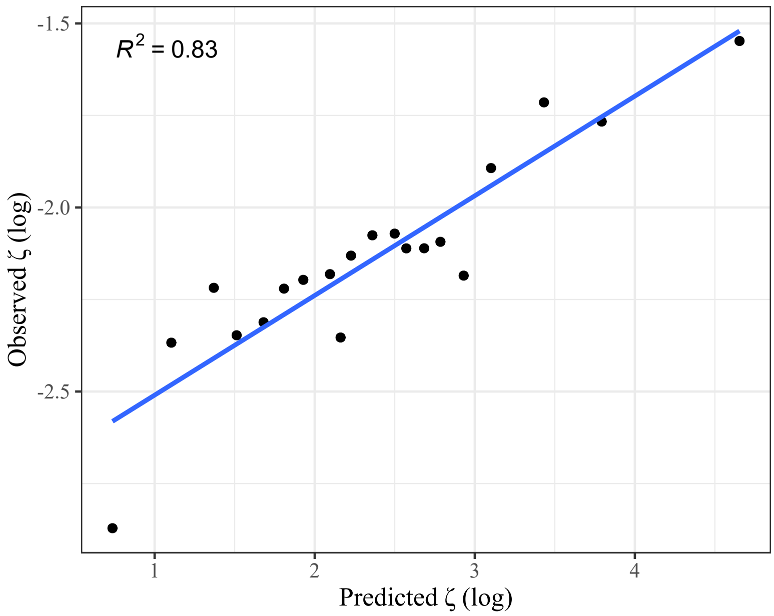
NOTES: Each dot corresponds to one of the 20 size bins. The x-axis gives the average size of firms in the bin (K_b). In panel (a), the y-axis is the share of firms that are larger than this size ($1 - F(K_b)$). In panel (b), the y-axis is the average squared geographical distance at which firms in the bin cite (Δ_b), in millions of kilometers. All citations over the period 1980-2017 are used. Innovator size is measured as the number of firms cited by the firm over that period. Distance is measured as the distance between the largest city of the countries of the citing and cited patents. Intranational citations and self-citations are excluded.

4.2 Testing the relation between ζ, λ and μ

Observed and predicted distance elasticity. To further test the validity of the model, one can test whether changes in λ or μ are systematically associated with changes of ζ as predicted by theory. We calculate $\hat{\zeta}$, $\hat{\lambda}$ and $\hat{\mu}$ on disaggregated samples at two different levels: country $\times$ year, and country $\times$ technology $\times$ decade. Since Chaney (2018) predicts that $\zeta = 1 + 2 \cdot \frac{\lambda - 1}{\mu}$, we can test whether this formula calculated with $\hat{\lambda}$ and $\hat{\mu}$ fits the observed $\hat{\zeta}$.

Figure 3 provides such piece of evidence, comparing the observed $\hat{\zeta}$ at the country $\times$ year level with the formula $1 + 2 \cdot \frac{\hat{\lambda}-1}{\hat{\mu}}$. Taken in logs, and with country fixed-effects, this comparison displays an excellent fit: there is a clear increasing relationship between both variables, with a R-squared of 0.83. The correspondance is however not perfect: $1 + 2 \cdot \frac{\hat{\lambda}-1}{\hat{\mu}}$ are typically much larger than $\hat{\zeta}$. This is not surprising, as applying this model to the context of trade also produces very large values of predicted ζ (Dewitte 2022). Predicted values are even larger in our context, given that departures for Zipf law are stronger than for exporters' size, while measured $\hat{\zeta}$ are systematically below unity. It is however very reassuring that, although the model may be insufficient to fully explain changes in $\hat{\zeta}$, there is a very strong statistical correspondence between changes in $\hat{\zeta}$ and changes in $1 + 2 \cdot \frac{\hat{\lambda}-1}{\hat{\mu}}$.

Figure 3: Link between observed ζ and ζ predicted by the model



NOTES: Based on $\hat{\lambda}$ and $\hat{\mu}$ estimated at the "year $\times$ country" level, we compute the predicted ζ as $1 + 2 \cdot \frac{\hat{\lambda}-1}{\hat{\mu}}$. We then relate these predicted ζ to their empirical counterparts (observed $\hat{\zeta}$), taking logs, introducing country fixed-effects, and grouping the data into 20 bins.

Decomposing the predictive power. An important dimension through which the context we study departs from the context of trade flows is that the number of sources new firms are endowed with, which we denote ι and measure as the intercept of the underlying μ regression, does not seem constant across countries and periods. This makes it an important dimension to control for, as for instance a lower μ (weaker relation between size and distance of citations) will have opposite interpretations if it is accompanied with an increase in ι (all firms cite further away, ζ should decrease), a stable ι (large firms cite closer, ζ should increase), or a decrease in ι (all firms, but large ones in particular, cite closer, and ζ should increase). Thus, to understand further what generates changes in $\hat{\zeta}$, we regress the obtained $\hat{\zeta}$ on $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\iota}$, and show the results in Table 4. Columns (1) and (2) are obtained at the country $\times$ year level (as Figure 3), while columns (3) and (4) correspond to the estimates at the country $\times$ technology $\times$ decade level.

Several striking facts emerge from Table 4. First, all results show an excellent fit of the model, with a global R-squared close to 90%, and a within R-squared (reflecting the share of the variance explained by the estimated parameters) around 55% on the country dimension, and 30% on the country–ipc dimension. This shows that changes in the parameters corresponding to the model predictions (λ and μ) drive changes in the measured effect of distance ζ. Second, variations in λ seem to have little influence on ζ: the estimated coefficient on λ is close to zero and hardly significant. Several reasons may explain this, and in particular the fact that λ shows rather limited variations across subsamples. Third, variations in μ are in contrast very sharply associated to changes in $\hat{\zeta}$, with coefficients that are the right sign, large and statistically very significant. Finally, the inclusion of ι seems important, as it is always very significant, and associated to a very stable coefficient across regressions.

Table 4: Predictive power : link between λ, μ and ζ

Observation unit	year-country		decade-country-ipc	
Model:	(1)	(2)	(3)	(4)
Variables				
μ	-2.33***	-2.28***	-1.24***	-1.32***
	(0.53)	(0.58)	(0.40)	(0.40)
λ	0.04**	0.00	-0.21	-0.22
	(0.02)	(0.07)	(0.14)	(0.15)
ι	-0.46***	-0.46***	-0.34***	-0.34***
	(0.07)	(0.11)	(0.09)	(0.08)
Fixed-effects				
Country	✓	✓	–	–
Period	–	✓	–	✓
Country-IPC	–	–	✓	✓
Fit statistics				
Observations	192	192	128	128
R^2	0.867	0.890	0.924	0.926
Within Adjusted R^2	0.564	0.596	0.321	0.318

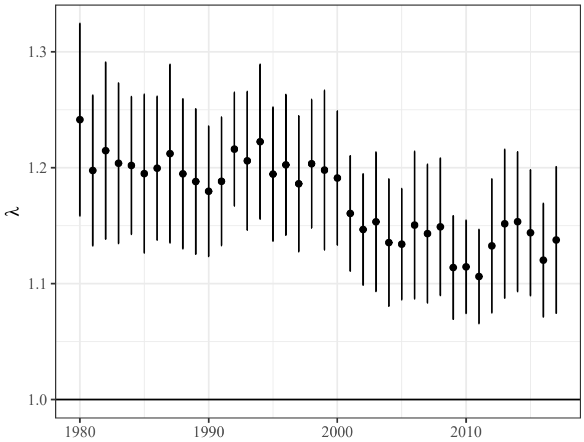
NOTES: $\hat{\mu}$, $\hat{\lambda}$, $\hat{\iota}$ and $\hat{\zeta}$ are estimated at the "year $\times$ country" level (columns (1) and (2)) or at the "decade $\times$ country $\times$ industry" (columns (3) and (4)), for the 10 largest countries (i.e. the 10 countries with the most outward citations). Observations for which $\hat{\mu}$ or $\hat{\zeta}$ are negative, or for which $\hat{\lambda}$ is below 1, are dropped from the sample. We then regress $\hat{\zeta}$ on $\hat{\mu}$, $\hat{\lambda}$ and $\hat{\zeta}$ with either country fixed-effects (columns (1) and (2)) or "decade $\times$ country" fixed-effects (columns (3) and (4)). In columns (2) and (4) we additionnally introduce period fixed-effects (year or decade).

Changes of the key parameters over the period. While one cannot directly link observed changes in ζ to the deep structural parameters dictating the arrival of new sources, the link between ζ, λ and μ is given by the model at any given point in time, such that changes in the former parameter over time can be related to changes of the latter parameters. In other words, now that we have established that changes in the model's theoretical predictions fit well the changes in the measured distance elasticity, one can study the variations of the model parameters and relate those to the variations of the measured distance elasticity.

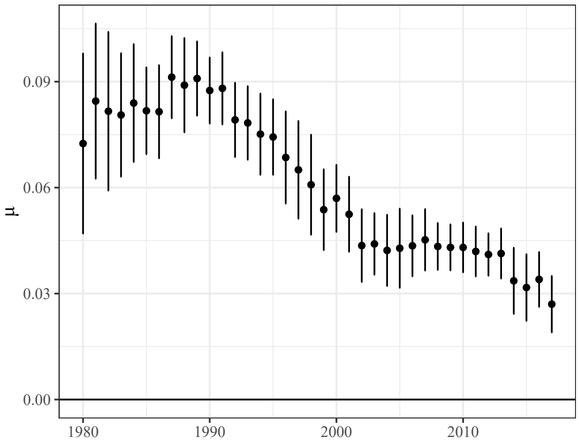
The underlying parameters λ, μ and ι, predicted by the network formation model, display large variations over the study period. This is illustrated in Figure 4, where we plot yearly estimates of $\hat{\lambda}$, $\hat{\mu}$ and $\hat{\iota}$. λ decreased slightly over time (Figure 4a), implying a growing concentration of innovation within large firms. Since smaller firms have less access to more distant knowledge sources, their reduced share in the firm sizes distribution implies, ceteris paribus, a lower distance elasticity of patent citations. In the meantime, those small firms became able to cite further away. This corresponds in our model to an increase in ι (Figure 4c). The generalization of long distance citations is likely to reflect the adoption of new information and communication technologies, such as the internet. However, this increase in the distance at

Figure 4: Estimates of λ, μ and ι over time.

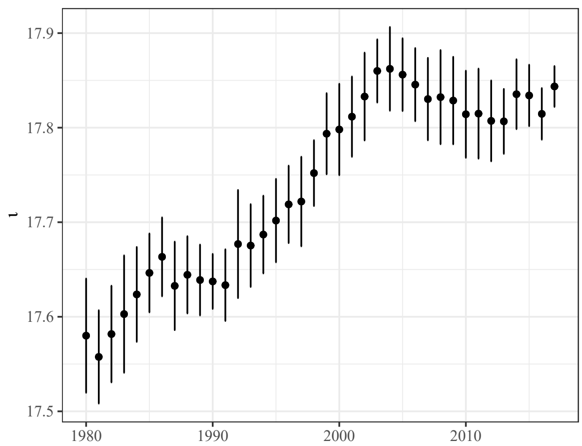
(a) Shape parameter of the Pareto distribution of innovator sizes (λ)



(b) Elasticity of the average squared distance of citations with respect to size (μ)



(c) ι over time



NOTES: λ, μ and ι are estimated from a series of cross-sectional regressions (respectively of equation (5) and (6)), one for each year. All patents are included in the sample. Innovator size is measured as the number of firms cited by the firm during the year. The distance is the geographical distance between the largest city of the countries of the citing and the cited patent. Standard errors are obtained using 100 bootstrap replications.

which small firms cite was partly compensated by the fading relation between firms' size and the distance at which they cite (μ decreases, as can be seen on Figure 4b). This suggests that the arrival of new sources decreased over time, which, along with an increased concentration of innovators, is in line with facts documented in the firm dynamics litterature (Akcigit and Ates 2021; 2023).

Overall, when comparing repeated cross sections, two parameters (λ and ι) are pushing towards a reduction of spatial frictions, while the third one (μ) corresponds to increased frictions. The combination of these facts is compatible with the stability of the distance elasticity documented as Fact #1: while the increase in the concentration of innovators implies that a larger share of citations is achieved by large firms with many distant sources, the decrease in μ implies that larger firms cite on average closer at the end of the period, a fact partly mitigated by the increase in the distance at which the smallest firms cite.

4.3 Composition effects and counterfactuals.

Over the study period, one might anticipate a decline in ζ owing to significant advancements in search technologies and expanded access to remote knowledge. This slight decrease in ζ corresponds to what we observe in the data, provided we account for the geographic recomposition of world innovation. Since the 1980s, new countries have emerged as key innovators, and these countries tend to have higher ζ, which compensates the decrease of ζ in initially large countries. Put differently, the decline expected within each unit was offset by the rising importance of units facing intrinsically higher search frictions. Once we correct for this composition effect, we find that the distance elasticity sligthly decreased between 1980 and 2017.

To analyze the geographic composition effects, we partition the world into 10 big countries (the countries with the most outward citations over the period), and a residual geographic aggregate ("rest of the world"). We define sample shares of each country i for each year t, w_{it}, as the share of the country in the total number of observations for that given year, considering the residual geographic aggregate as one single country. The distance effect in year t, ζ_t can be recovered as a weighted average of the country specific distance coefficients:[14]

$$\zeta_t^r = \sum_{i \in C} w_{it} \cdot \hat{\zeta}_{it} \tag{7}$$

Such a decomposition is useful because it allows to compute a "counterfactual" predicted distance elasticity, setting sample shares to their 1980s level: $\zeta_t^{\text{init}} = \sum_{i \in C} w_{i,1980s} \cdot \hat{\zeta}_{it}$. This corresponds to the distance elasticity that would have been observed in t, had the geographic or industry distribution of outward citations remained steady since the 1980s.

The composition effect is the difference between the predicted distance elasticities with actual sample shares and with 1980s shares: $\Delta\zeta_t^{\text{init}} = \sum_{i \in C}(w_{i,1980s} - w_{it}) \cdot \hat{\zeta}_{it}$. The counterfactual distance elasticity, purged from composition effects, is then obtained as:

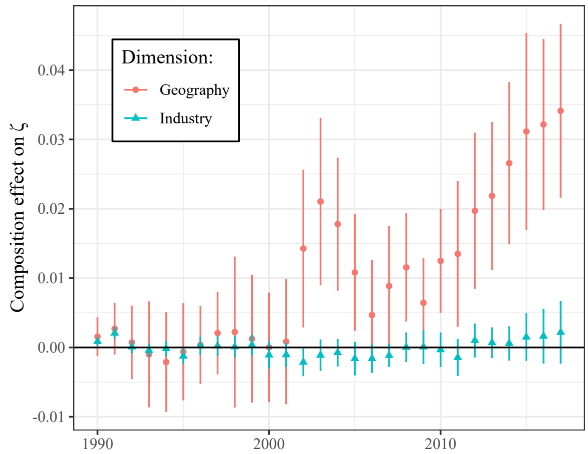
$$\zeta_t^{\text{ctrf}} = \hat{\zeta}_t + \Delta\zeta_t^{\text{init}} = \hat{\zeta}_t + \sum_{i \in C}(w_{i,1980s} - w_{it}) \cdot \hat{\zeta}_{it} \tag{8}$$

Figure 5 uses respectively shares of countries and technologies (IPC 1-digit) fixed in the 1980s, and recomputes an aggregate ζ using these shares. Figure 5a measures the deviation of this reaggregated ζ (with shares initially fixed) to the observed ζ. It shows that the technolog-

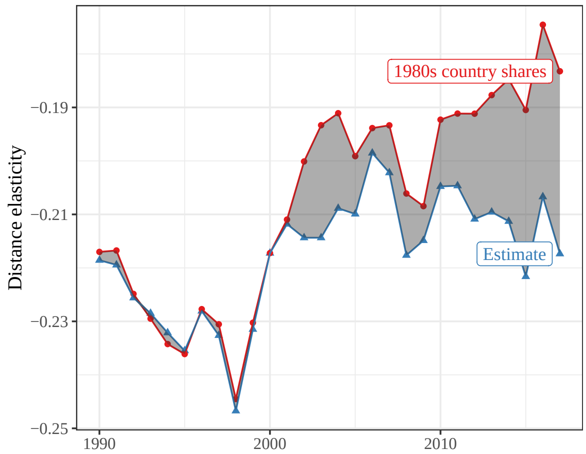
14. Figure C2 in the Appendix shows that the ζ_t^r obtained following this procedure are very close to the estimated $\hat{\zeta}_t$, confirming the validity of this method.

Figure 5: Composition effects underlying changes in ζ and counterfactual

(a) Geographic and industry composition effect for ζ over time



(b) Counterfactual ζ, keeping country shares to their 1980s level



NOTES: To obtain the country x year distance elasticities, we estimate $Y_{tijk} = \exp\left(\text{FE}_{tik} + \text{FE}_{tjk} + \sum_{s,c} \zeta_{sc}\, \text{dist}_{ij} \times \mathbb{1}(t = s, i = c) + \varepsilon_{tijk}\right)$, where c denotes either the country itself if the country is big, or a residual geographic aggregate ("rest of the world") if the country is small. In figure a), we plot the composition effects, setting the country sample shares ("Geography") or the IPC 1-digit technology class sample shares ("Industry") to their 1980s level. Bars display the 95% confidence intervals. In figure b), we plot both the ζ directly estimated using equation (1) (blue curve), and a counterfactual ζ, that would have been observed had the country composition of the sample remained to its 1980s levels (red curve). The composition effects and the counterfactual ζ are obtained from equation (8)

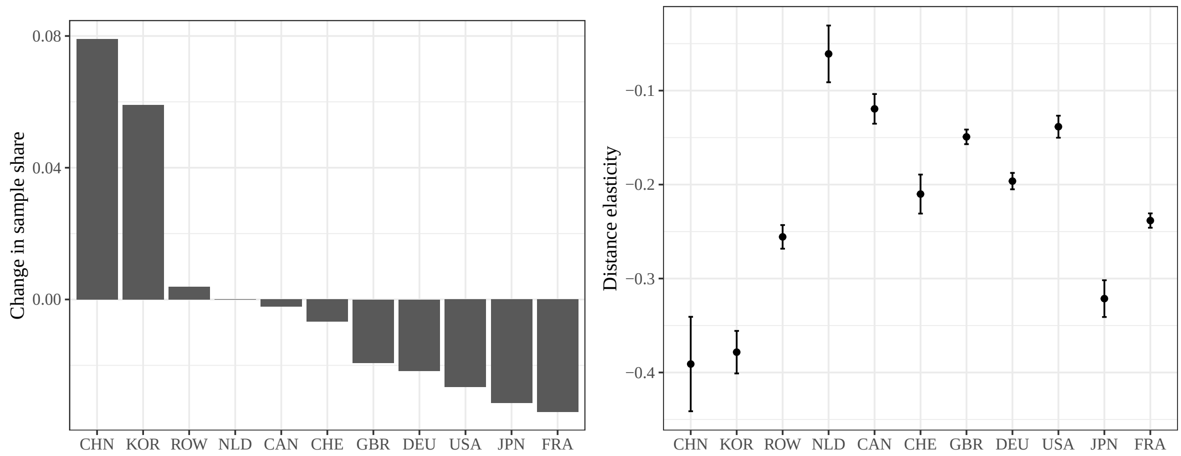
ical recomposition over the period, while it may have been quantitatively important, had little impact on the overall effect of distance. The geographic composition, however, had important consequences on the overall distance elasticity: it significantly increased the measured distance elasticity. To account for the change in country composition, we compute for each year a counterfactual ζ, which would have been observed had the geographic composition of the sample remained to its initial levels. This is illustrated through Figure 5b, which shows the actual estimates as a blue line, and the counterfactual estimates, had the share of countries remained the same as in the 1980s, as a red line.

We further explore the link between the observed stability of ζ and the underlying country composition. Figure 6 shows both the change in the sample share of each main country over the period (Figure 6a), and the associated ζ over the period of these countries (Figure 6b). A striking fact emerges from this figure: China and Korea's shares rose considerably over the period, jointly representing a 14 pp increase in the share of total observations they represent. Conversely, the US decreased a bit ($\approx$ -2.5 pp), Japan decreased as well ($\approx$ -3 pp), while Europe overall decreased a lot (around -8 pp). This surge of China and Korea in global innovation pushes ζ to increase, as these countries have an associated distance elasticity which is very large in absolute terms (around -0.4, almost double the aggregate elasticity measured).

Figure 6: Country-level changes in sample share and associated distance elasticities

(a) Change in the country share in total number of observations

(b) ζ estimated over the period 1980-2017, by country



NOTES: Figure a) Change in sample shares correspond to the difference between the country share in the number of observations in 2017 and its share in the number of observations during the decade 1980-1989. Figure b) To obtain the country x year distance elasticities, we estimate $Y_{tijk} = \exp\left(\text{FE}_{tik} + \text{FE}_{tjk} + \sum_{s,c} \zeta_{sc}\, \text{dist}_{ij} \times \mathbb{1}(t = s, i = c) + \varepsilon_{tijk}\right)$, where c denotes either the country itself if the country is big, or a residual geographic aggregate ("rest of the world") if the country is small.

While it is difficult to link directly this geographic composition effect to underlying parameters reflecting different patterns of link formation, Figure C1 in the Appendix provides country by country parameters, estimated over the whole period of study. The figure shows facts which are broadly consistent with what different distance elasticities suggest. In particular, Korea has one of the highest λ, while China has an extremely low μ. An important point to bear in mind is however that, while the estimation of ζ neutralizes the geographic composition of patents that could be cited, through origin and destination country fixed-effects, the estimation of underlying parameters does not. This limits the comparability of country estimates as the geographic distribution of potentially cited patents may considerably vary.

A remaining question is why East Asian countries display such a high ζ, reflecting considerable search frictions. The case of Japan, which was already a very large innovative country at the beginning of the period but displays a distance elasticity of citations which is comparable to that of China and Korea, suggests that there might be a common factor. The literature on trade has shown the crucial role of language differences in explaining geographic frictions (Melitz and Toubal 2014; Egger and Lassmann 2012), which have been shown to matter for flows of ideas as well (Keller 2002). Language difference is therefore a natural candidate to explain frictions in knowledge diffusion. In particular, ideogrammatic native languages in these countries may impose larger search costs on innovators, in an initial equilibrium where the dominant language for innovation is English. Indeed, distance elasticities are particularly low in English-speaking countries such as the US, Great Britain and Canada, and are midrange in non-english speaking countries with languages in latin alphabet.

5 Conclusion

This paper shows that the negative effect attributed to distance on international knowledge flows can be explained by the spatial pattern in the dynamics of network formation between innovators.

We start by uncovering two stylized facts. First, at the aggregate level, the negative relation between citation flows and geographical distance has remained constant over the past four decades, in spite of tremendous progress in information and communication technologies. Second, at the micro level, a firm's existing sources have a double influence on future flows of citations: they make more likely both citations toward never cited patents of these sources, and towards patents cited by these sources.

Based on these findings, we use and extend the network formation model developed by Chaney (2018) to draw aggregate implications from the above phenomenon. The theoretical aggregate predictions of the model hold remarkably well in the data. The sizes of innovators—measured as the number of cited firms—are Pareto-distributed (and even Zipf distributed), and the average squared distance at which innovators cite is an increasing power function of their size. The Zipf distribution of innovator sizes, as well as the systematic increasing relationship between size and distance at which firms are able to cite, are novel findings. They allow generating a negative effect of distance on aggregate citation flows: if small firms are far more numerous than big ones and if they cite relatively closer, flows will naturally vary negatively with distance.

Lastly, we measure distance elasticities at disaggregated levels and run counterfactuals fixing countries or technologies shares at the beginning of the period. We find that, while changes in the technological composition had no impact on the aggregate effect of distance, changes in the composition of countries are crucial to explain the stability of the distance elasticity observed over the period. In particular, we show that the effect of distance should have decreased since the 1980s, but the rise of East Asia, whose countries display very large distance elasticities, offset the overall gains. We hypothesize that language barriers, which are known to have a sizable impact on trade flows, also persistently affect knowledge flows.

References

Aghion, Philippe, Ufuk Akcigit, and Peter Howitt. 2015. "Lessons from Schumpeterian growth theory." *American Economic Review* 105 (5): 94–99.

Aghion, Philippe, and Xavier Jaravel. 2015. "Knowledge Spillovers, Innovation and Growth." *Economic Journal,* no. 583, 533–573.

Agrawal, Ajay, Devesh Kapur, and John McHale. 2008. "How do spatial and social proximity influence knowledge flows? Evidence from patent data." *Journal of Urban Economics* 64 (2): 258–269.

Akcigit, Ufuk, and Sina T Ates. 2021. "Ten Facts on Declining Business Dynamism and Lessons from Endogenous Growth Theory." *American Economic Journal: Macroeconomics* 13 (1): 257–98.

———. 2023. "What happened to US business dynamism?" *Journal of Political Economy* 131 (8): 2059–2124.

Akers, Noël J. 2000. "The referencing of prior art documents in European patents and applications." *World Patent Information* 22 (4): 309–315.

Albert, R., and A.-L. Barabási. 2002. "Statistical mechanics of complex networks." *Reviews of Modern Physics* 74:47–97.

Alcácer, Juan, and Michelle Gittelman. 2006. "Patent Citations as a Measure of Knowledge Flows: The Influence of Examiner Citations." *Review of Economics and Statistics* 88 (4): 774–779.

Almeida, Paul, and Bruce Kogut. 1997. "The Exploration of Technological Diversity and the Geographic Localization of Innovation." *Small Business Economics* 9 (1): 21–31.

Axtell, Robert L. 2001. "Zipf distribution of US firm sizes." *Science* 293 (5536): 1818.

Bloom, Nicholas, Charles Jones, John van Reenen, and Michael Webb. 2020. "Are Ideas Getting Harder to Find?" *American Economic Review* 110 (4): 1104–44.

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

Breschi, Stefano, and Francesco Lissoni. 2009. "Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows." *Journal of economic geography* 9 (4): 439–468.

Buera, Francisco J., and Ezra Oberfield. 2020. "The Global Diffusion of Ideas." *Econometrica* 88 (1): 83–114.

Chaney, Thomas. 2014. "The network structure of international trade." *American Economic Review* 104 (11): 3600–3634.

———. 2018. "The Gravity Equation in International Trade: An Explanation." *Journal of Political Economy* 126 (1): 150–177.

Corsino, Marco, Myriam Mariani, and Salvatore Torrisi. 2019. "Firm strategic behavior and the measurement of knowledge flows with patent citations." *Strategic Management Journal* 40 (7): 1040–1069.

Cotropia, Christopher A, Mark A Lemley, and Bhaven Sampat. 2013. "Do applicant patent citations matter?" *Research Policy* 42 (4): 844–854.

Dewitte, Ruben. 2022. "The gravity equation in international trade: A note." *Journal of Political Economy* 130 (5): 1412–1418.

Duguet, Emmanuel, and Megan MacGarvie. 2005. "How well do patent citations measure flows of technology? Evidence from French innovation surveys." *Economics of Innovation and New Technology* 14 (5): 375–393.

Egger, Peter H, and Andrea Lassmann. 2012. "The language effect in international trade: A meta-analysis." *Economics Letters* 116 (2): 221–224.

EPO. 2016. *Guidelines for Examination, Part B (Guidelines for Search).* Technical report. European Patent Office.

Gabaix, Xavier. 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30 (1): 185–206.

Ghiglino, Christian. 2012. "Random walk to innovation: Why productivity follows a power law." *Journal of Economic Theory* 147 (2): 713–737.

Gibrat, Robert. 1931. *Les inégalités économiques: applications: aux inégalités des richesses, à la concentration des entreprises, aux populations des villes, aux statistiques des familles, etc: d'une loi nouvelle: la loi de l'effet proportionnel.* Librairie du Recueil Sirey.

Griffith, Rachel, Sokbae Lee, and John Van Reenen. 2011. "Is distance dying at last? Falling home bias in fixed-effects models of patent citations." *Quantitative economics* 2 (2): 211–249.

Head, Keith, Yao Amber Li, and Asier Minondo. 2019. "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics." *Review of Economics and Statistics* 101 (4): 713–727.

Head, Keith, and Thierry Mayer. 2014. "Gravity Equations: Workhorse, Toolkit, and Cookbook." Chap. 3 in *the Handbook of International Economics,* 4:131–195. Elsevier.

Jackson, Matthew, and Brian Rogers. 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review* 97 (3): 890–915.

Jaffe, Adam B., Manuel Trajtenberg, and Michael S. Fogarty. 2000. "Knowledge Spillovers and Patent Citations: Evidence from a Survey of Inventors." *American Economic Review, Papers and Proceedings* 90 (2): 215–218.

Jaffe, Adam B., Manuel Trajtenberg, and Rebecca Henderson. 1993. "Geographic localization of knowledge spillovers as evidenced by patent citations." *Quarterly Journal of Economics,* 577–598.

Keller, Wolfgang. 2002. "Geographic Localization of International Technology Diffusion." *American Economic Review* 92 (1): 120–142.

Kerr, William R. 2008. "Ethnic scientific communities and international technology diffusion." *Review of Economics and Statistics* 90 (3): 518–537.

Kuhn, Jeffrey, Kenneth Younge, and Alan Marco. 2020. "Patent citations reexamined." *The RAND Journal of Economics* 51 (1): 109–132.

Lampe, Ryan. 2012. "Strategic citation." *Review of Economics and Statistics* 94 (1): 320–333.

Lei, Zhen, and Brian D. Wright. 2017. "Why weak patents? Testing the examiner ignorance hypothesis." *Journal of Public Economics* 148:43–56.

Li, Yao Amber. 2014. "Borders and distance in knowledge spillovers: Dying over time or dying with age?—Evidence from patent citations." *European Economic Review* 71:152–172.

Maurseth, Per Botolf, and Bart Verspagen. 2002. "Knowledge Spillovers in Europe: A Patent Citations Analysis." *Scandinavian Journal of Economics* 104 (4): 531–45.

Mayer, Thierry, and Soledad Zignago. 2011. *Notes on CEPII's distances measures: The GeoDist database.* Working Papers. CEPII research center.

Melitz, Jacques, and Farid Toubal. 2014. "Native language, spoken language, translation and trade." *Journal of International Economics* 93 (2): 351–363.

Murata, Yasusada, Ryo Nakajima, Ryosuke Okamoto, and Ryuichi Tamura. 2014. "Localized Knowledge Spillovers and Patent Citations: A Distance-Based Approach." *Review of Economics and Statistics* 96 (5): 967–985.

Nigai, Sergey. 2017. "A tale of two tails: Productivity distribution and the gains from trade." *Journal of International Economics* 104:44–62.

Peri, Giovanni. 2005. "Determinants of knowledge flows and their effect on innovation." *Review of Economics and Statistics* 87 (2): 308–322.

Singh, Jasjit. 2005. "Collaborative Networks as Determinants of Knowledge Diffusion Patterns." *Management Science* 51 (5): 756–770.

Thompson, Peter. 2006. "Patent Citations and the Geography of Knowledge Spillovers: Evidence from Inventor- and Examiner-added Citations." *Review of Economics and Statistics* 88 (2): 383–388.

Thompson, Peter, and Melanie Fox-Kean. 2005. "Patent citations and the geography of knowledge spillovers: A reassessment." *American Economic Review* 95 (1): 450–460.

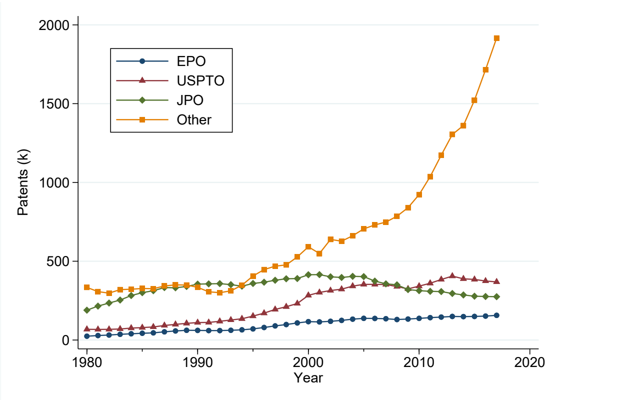
ONLINE APPENDIX

A Data Appendix

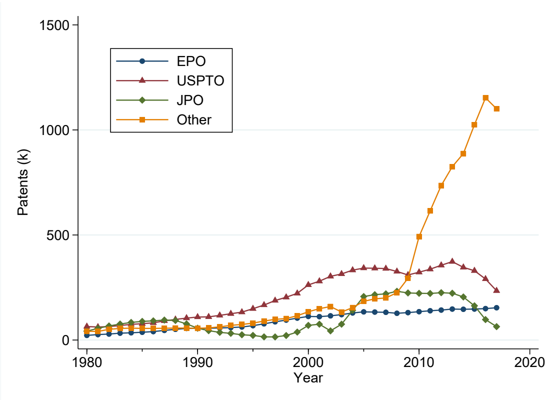
A.1 Description of the Patstat database

Figure A1: Number of patents/citations, decomposed by patent office

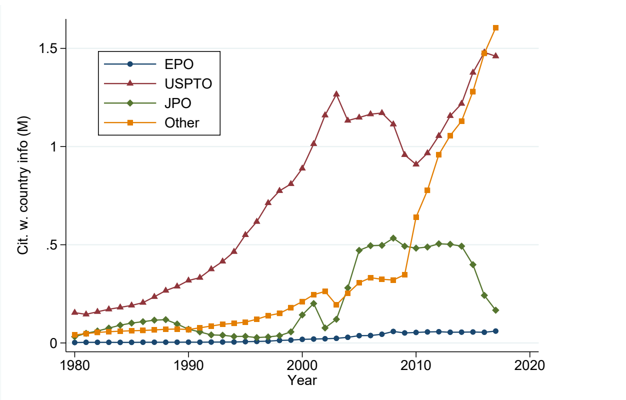
(a) All patents



(b) Patents with citation info



(c) Number of citations



(d) Applicant added citations

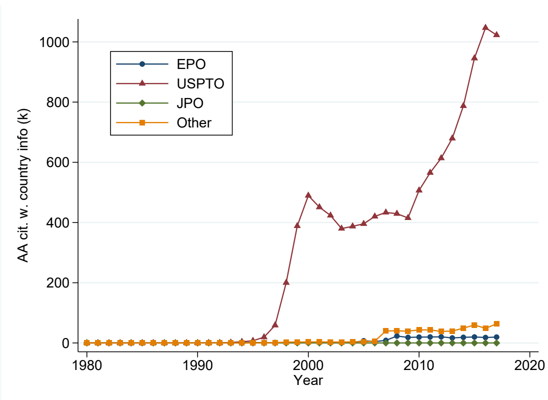


Figure A2: Patents/citations for which we have geographic information (country)

(a) Patents **(b)** Citations

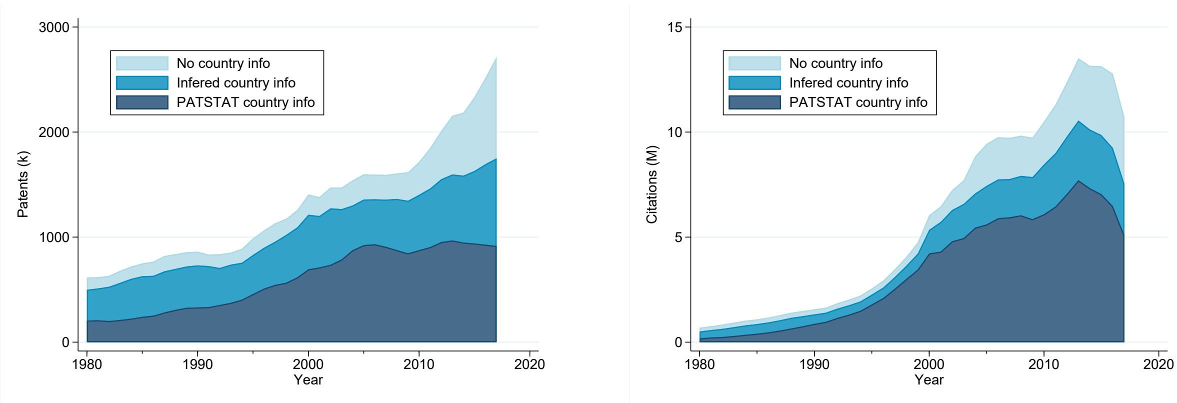
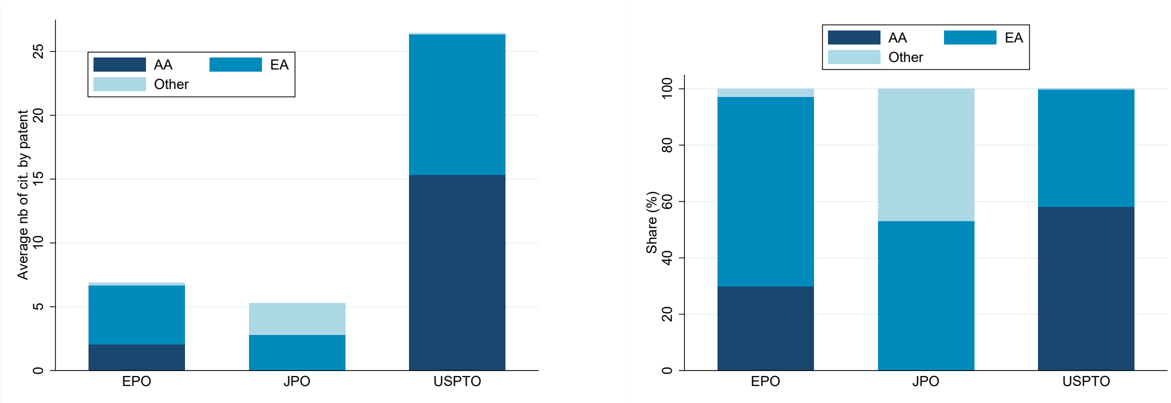


Figure A3: Type of outward citations, by patent office

(a) Average number per patent **(b)** Proportion



NOTES:Patents with priority year posterior to 2000

A.2 Patent applicants and country.

Patent applications distinguish between the people who actually developed the claimed invention (called the inventors) and those who will obtain the legal rights over the invention if the application is successful (equivalently called the applicants or the assignees throughout this paper). Notably, inventors are usually employees of the institution which obtains the legal rights over the invention. Therefore, inventors are always private individuals, while the vast majority of assignees are firms. Since our focus is on firms, we determine the country of a patent through the country of its assignee. However, 11% of the applications have several assignees, potentially based in different countries. In such case, we consider the application to be located in the country that appears most frequently among the assignees (the mode), and if there is no mode, we assign randomly one of the assignees' countries to the patent.

Patents do not include unique firm identifiers, therefore the allocation of a patent to a firm can be made only through the assignee name indicated on the patent. A common issue is that the applicant's name may be different even for patents belonging to the same firm due to spelling mistakes, spelling variations, and national units of large companies. Therefore, some algorithms were developed to harmonize applicant names. Patstat contains several name harmonizations, of which we use the Patstat Standardized Name (PSN) applicant identifier.[15] Note also that along with name harmonization, Patstat contains information on the type (firm, university, etc.) of each applicant. Unless specified otherwise, we keep only the applicants that are signalled as firms in this harmonization.

We use a simple procedure to improve information on the country of the applicants: we carry the information forward based on the name harmonization work performed by Patstat. Suppose the country is missing for a patent, but is available for another patent granted to the same assignee: we consider that the country of the former patent is also the one of the latter patent. Thanks to this method, we infer geographic information for an additional third of the patents, which leaves us with only a small share of patents without country information, as illustrated in Figure A2.

A.3 Observable characteristics of applicant and examiner citations.

The observable characteristics we study are defined in the following way.

- **Age:** Age is simply the difference between the priority date of the citing patent and the priority date of the cited patent.

- **Quality:** We build a proxy for the quality of each patent by regressing the number of citations this patent received on a set of fixed-effects absorbing the effects of technological classes (IPC 3 digits), priority year and office.[16] In order to use log-transformed values in the regressions, we shift all values by the absolute value of the minimum.

- **Geographical Distance:** Spatial distance is determined based on the cities of the assignees of the citing and cited patents. In the case where there are several applicants located in different cities, we take the mode of the different cities, or we randomly choose the city of one of the applicants if there is no mode.

- **Technological Distance:** Additionally to the previous variables, we build a measure of technological distance between the citing and the cited patents based on the IPC classes

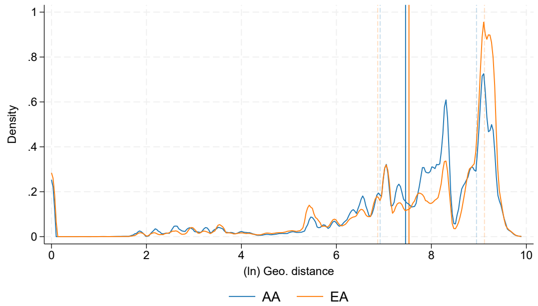
15. Provided by ECOOM https://www.ecoom.be/en/EEE-PPAT it is automated and is particularly accurate for the largest patentees, which is crucial when estimating a size distribution. Moreover, it is available for assignees at all offices represented in Patstat, while the HAN harmonization conducted by the OECD is mostly for the EPO.

16. To include IPC 3 digits fixed-effects, we need to assign a single IPC3 digit of each patent (a patent may belong to several IPC 3 digits, whereas our strategy requires that each patent is associated with one single IPC3 digit). To determine the main IPC 3 digit of a patent, we consider all the IPC 6 digits of this patent, each of which corresponding to a single IPC 3 digits, and find the mode of IPC 3 digit based on this.

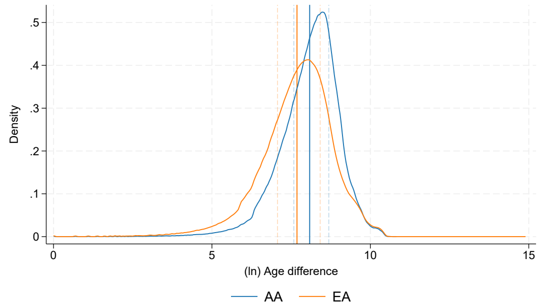
in which it has been filed. This measure was introduced in Bloom, Schankerman, and Reenen (2013), and refines approaches used in Jaffe, Trajtenberg, and Henderson (1993), Thompson and Fox-Kean (2005) and Murata et al. (2014).

Figure A4: Distribution of observable characteristics in applicant-added and examiner-added citations

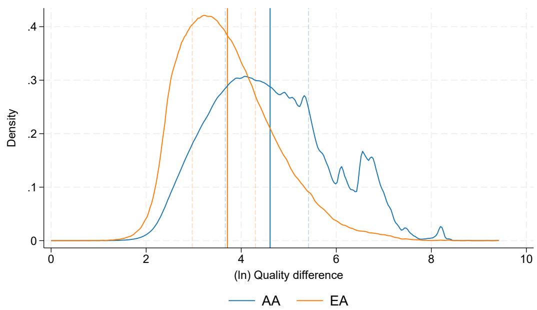
(a) log(Geographical distance)



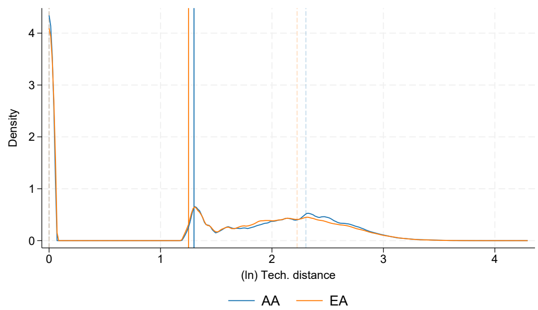
(b) log(Age difference)



(c) Quality gap (dlog)



(d) log(Technological distance)



(a) Upper left panel: geographical distance between the citing and the cited patent. (b) Upper right panel: age of the cited patent at the time of the citation. (c) Lower left panel: quality difference between cited and citing patent. (d) Lower right panel: Mahalanobis technological distance between the citing and the cited patent. Blue lines represent applicant added citations (distribution, mean, 1st and 3rd quartile), orange lines represent examiner added citations.

A.4 Description of examiner citations

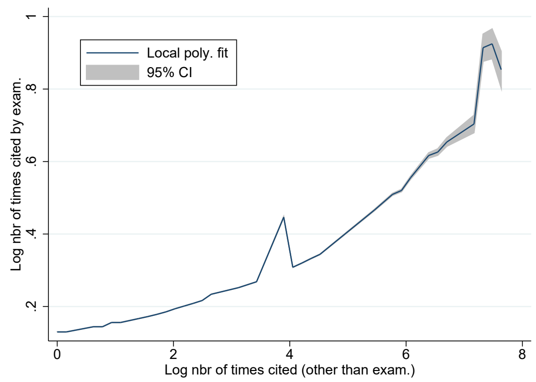
This section examines the PatEX database from USPTO's Public PAIR data, which records information about the examination process at the USPTO, matched with our sample of USPTO patent applications from Patstat. It reveals the following findings.

Time spent on a patent application by an examiner is substantial: after dropping very occasional examiners (the ones with less than 5 applications), the average examiner handles 40 patent applications per year, with the 95th percentile being slightly above 100, meaning that even very busy examiners deal with two applications in an average working week. This suggests that the citations added in the process of examination should have been cautiously analyzed. Similarly, examination is conducted by one person only.

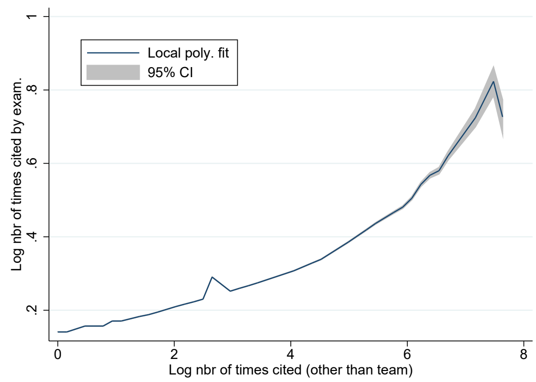
Examiners appear to be very specialized in their field: keeping only the eight technological centers as they exist today (to avoid counting organizational changes as movements), 78% of examiners remained their whole career in one of the centers, while 86% of examiners handled patents for less than 4 of the 589 technological divisions called art units[17] over their career.

Figure A5: Correspondence between the number of times an examiner cites a patent and the number of times other examiners cite it.

(a) Number of times cited by an examiner vs by other examiners



(b) Number of times cited by an examiner vs by other examiners outside of her team



There appears to be limited habit formation in examiners' behavior. As Figure A5 shows, patents cited several times by the same examiner also tend to be cited many times by other examiners, even when we consider only the ones outside the examiner's art unit (to exclude potential peer-effects). Looking at the technological distance between the patent application assessed by the examiner and the patents she cites, as shown in Table A1, we find that the first time an examiner cites a patent, the technological distance is only 1% of a standard deviation lower, or equivalently that each additional time a patent is cited by a given examiner implies an average increase of technological distance of .4% of a standard deviation. This means that, while habit formation in the way examiners cite may exist, it implies very small losses in the accuracy of citations as evidenced by our measure of technological distance.

17. Art units are grouped generally by 10 into clusters which include fields such as "Memory access and control", "Digital and optical communications", "Immunology, Receptor/Ligands, Cytokines Recombinant Hormones, and Molecular Biology", etc.

Table A1: Technological distance in multiple citations by examiners

	(1) Log tech dist	(2) Log tech dist
First citation by examiner	-0.012***	
	(0.001)	
Rank of examiner citation		0.004***
		(0.000)
Examiner FE	Yes	Yes
Dest. Pat. FE	Yes	Yes
Nbr of obs	9.593e+06	9.593e+06
R-sq	0.556	0.556

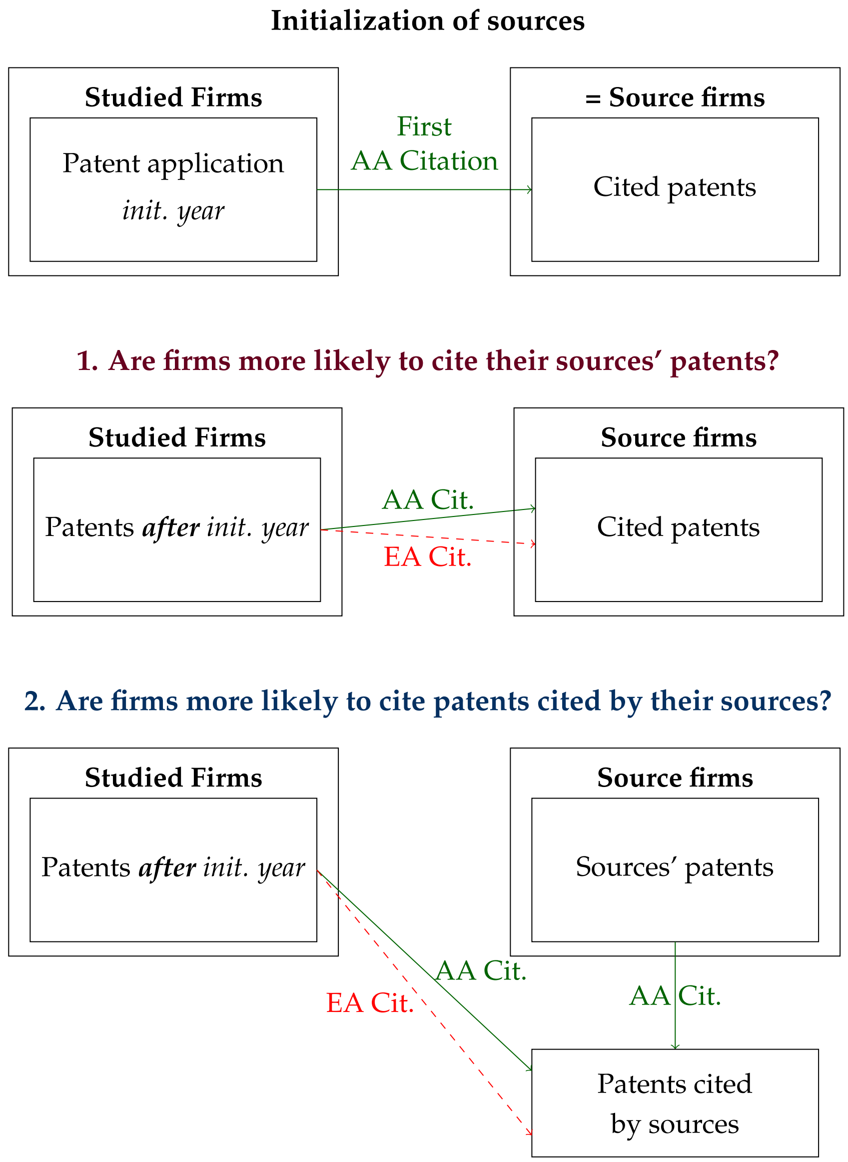
NOTES: The sample is composed of all citations to destination patents cited more than once by the same USPTO examiner. The dependent variable is the standardized technological distance between the citing and the cited patent (Mahalanobis distance calculated on IPCs 3 digits). "First citation by examiner" is a dummy variable taking value 1 when a patent is cited for the first time by an examiner. "Rank of examiner citation" is a variable taking value n when a citation corresponds to the n^{th} time an examiner cites a patent. Standard errors are clustered at the examiner level. *** pvalue < .01, ** pvalue < 0.05, * pvalue < 0.1.

B Additional tables and figures on stylized facts

Table B1: Summary statistics on the estimation sample for Fact #2

	Applicant-added citations				Examiner-added citations			
	mean	sd	p10	p90	mean	sd	p10	p90
Contact	0.190	0.392	0.000	1.000	0.050	0.218	0.000	0.000
Cited by Contact	0.324	0.468	0.000	1.000	0.082	0.275	0.000	0.000
Cited by Contact before initialization year	0.288	0.453	0.000	1.000	0.097	0.296	0.000	0.000
Firm already cited by applicant	0.176	0.380	0.000	1.000	0.102	0.303	0.000	1.000
Firm already cited	0.650	0.477	0.000	1.000	0.588	0.492	0.000	1.000
Patent already cited by applicant	0.287	0.452	0.000	1.000	0.068	0.252	0.000	0.000
Already cited before initialization year	0.284	0.451	0.000	1.000	0.088	0.284	0.000	0.000
Patent family already cited	0.388	0.487	0.000	1.000	0.127	0.333	0.000	1.000
Ln(Age Diff.)	8.085	0.922	6.933	9.098	7.684	1.127	6.303	8.962
Ln(Quality Diff.)	4.608	1.244	3.050	6.500	3.710	0.987	2.563	5.055
Ln(Tech. Dist.)	1.300	1.109	0.000	2.644	1.252	1.096	0.000	2.611
Ln(Geo. Dist.)	7.457	2.085	5.021	9.218	7.529	2.260	4.872	9.283
Nb. of citing firms	3.3e+05	0.000	3.3e+05	3.3e+05	4.5e+05	0.000	4.5e+05	4.5e+05
Nb. of citing patents	1.3e+06	0.000	1.3e+06	1.3e+06	2.1e+06	0.000	2.1e+06	2.1e+06
Observations	41424949				24825769			

Figure B1: Design of the tests



NOTES: AA Citation: Citation added by the applicant; EA Citation: Citation added by an examiner. The set of studied firms is made of a randomly picked third of all firms having patented both in the initialization year and in any subsequent year.

Table B2: Additional robustness checks on influence of existing links on citations

Dep. var.: Patent cited by the applicant	Baseline	Contact size		Time window		Cited patent FE	
		$\leq$ p99.9	$\leq$ p99.99	$[i; i+3]$	$[i; i+5]$		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Source	0.020***	0.018***	0.020***	0.025***	0.023***	0.023***	0.021***
	(0.0002)	(0.0002)	(0.0003)	(0.0004)	(0.0003)	(0.0003)	(0.0003)
Cited by Source	0.014***	0.013***	0.013***	0.016***	0.016***	0.005***	0.004***
	(0.0003)	(0.0003)	(0.0003)	(0.0005)	(0.0004)	(0.0003)	(0.0003)
Mean of the dep. variable	0.626	0.634	0.606	0.605	0.605	0.626	0.626
Number of citing firms	461.4k	454.2k	447.1k	286.8k	364.1k	461.4k	461.4k
Number of citing patents	2.2M	2.1M	2.1M	1M	1.5M	2.2M	2.2M
Number of observations	59.4M	51.6M	44.5M	18.6M	29.1M	63.8M	58.7M
Citing patent × cohort FE	✓	✓	✓	✓	✓	✓	✓
Cited firm FE	✓	✓	✓	✓	✓	–	–
Cited patent FE	–	–	–	–	–	✓	✓
Pairwise controls	✓	✓	✓	✓	✓	–	✓
Past citations controls	✓	✓	✓	✓	✓	✓	✓

NOTES: This table reports the coefficients corresponding to the specification described in Fact #2 (section 2.3), conducting several robustness checks. This table presents regression results obtained from estimating specification (2) through ordinary least squares: the coefficient labelled "Source" corresponds to β_1, the coefficient "Cited by Source" to β_2. Standard-errors are clustered at the "citing patent × cohort" level. The estimation sample contains all patent citations from a randomly selected third of patent applicants in each given year between 2000 and 2015. The dependent variable is a dummy variable indicating if the citation was added by the applicant or not. Column (1) is our baseline estimate, columns 2 and 3 replace the maximal size percentile for which firms are considered as potential sources, from p99 (removing top 1%) in the baseline to p99.9 (removing top 0.1%) or p99.99 (removing top 0.01%). Columns 4 and 5 impose a time window between the year in which sources are initialized and the year of application of the subsequent citing patent, of either 3 years max (column 4) or 5 years max (column 5). Columns 6 and 7 show coefficients similar to the baseline, but including cited patent rather than cited applicant fixed-effects.

Table B3: Influence of existing links on citations, using conditional logit estimations

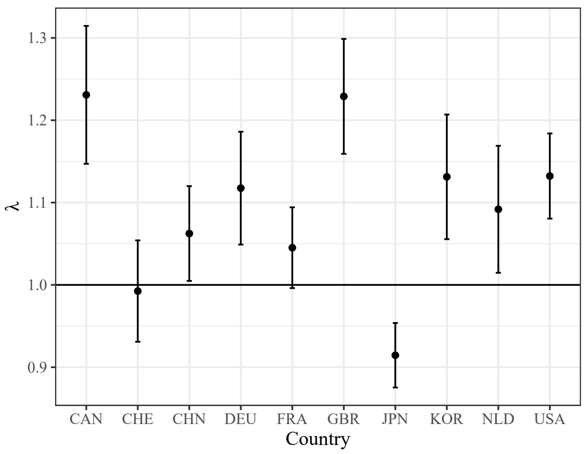
Dep. var.: Patent cited by the applicant			
	(1)	(2)	(3)
Source	1.343***	1.281***	1.259***
	(0.0027)	(0.0027)	(0.0028)
Cited by Source	1.351***	1.212***	1.135***
	(0.0028)	(0.0028)	(0.0028)
Mean of the dep. variable	0.566	0.566	0.566
Number of citing firms	299k	299k	299k
Number of citing patents	1.1M	1.1M	1.1M
Number of observations	27.6M	27.6M	25.2M
Citing patent FE	✓	✓	✓
Pairwise controls	–	–	✓
Past citations controls	–	✓	✓

NOTES: This table reports the coefficients corresponding to the specification described in Fact #2 (section 2.3). This table reports the coefficients corresponding to the specification described in Fact #2 (section 2.3), conducting several robustness checks. Coefficients are obtained through a conditional logit regression, at the level of citing patents, and are exponentiated to be interpretable as odds-ratios. The coefficient labelled "Source" corresponds to β_1, the coefficient "Cited by Source" to β_2. Standard-errors are clustered at the "citing patent $\times$ cohort" level. The estimation sample contains all patent citations from a randomly selected third of patent applicants in each given year between 2000 and 2015. The dependent variable is a dummy variable indicating if the citation was added by the applicant or not.

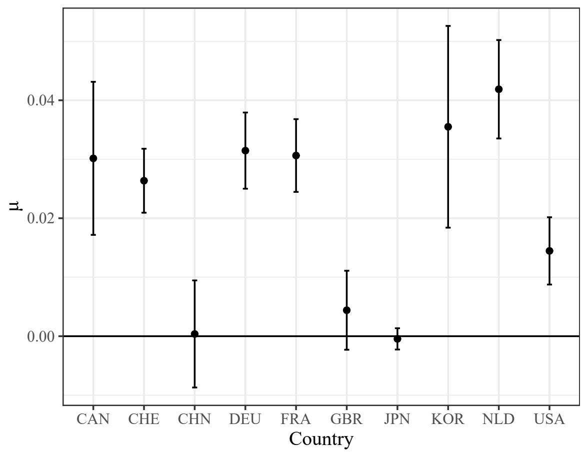
C Additional tables and figures on estimation

Figure C1: Parameters λ, μ and ι estimated by country, over the whole period (1980-2017).

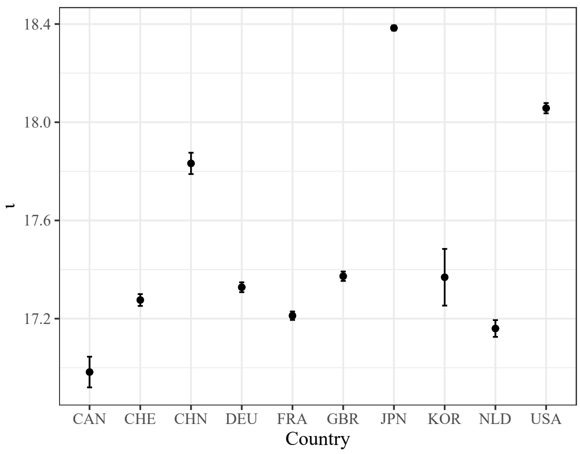
(a) λ by country



(b) μ estimated by country

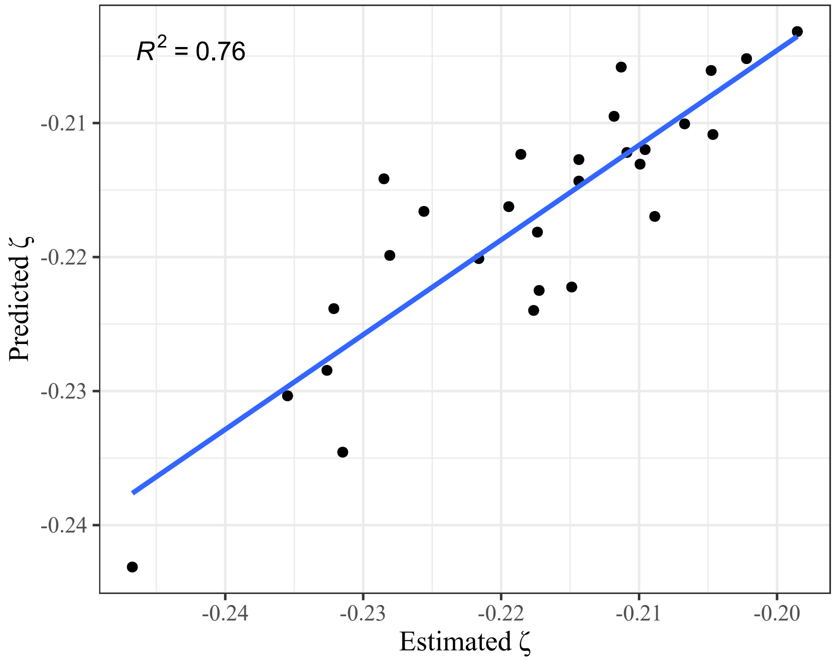


(c) ι by country



NOTES: λ, μ and ι are estimated from a series of regressions (respectively of equation (5) and (6)), one for each country. All patents are included in the sample. Innovator size is measured as the number of firms cited by the firm during the period 1980-2017. The distance is the geographical distance between the largest city of the countries of the citing and the cited patent. Standard errors are obtained using 100 bootstrap replications.

Figure C2: Link between ζ directly estimated at the year level and ζ recomposed from country $\times$ year estimates



NOTES: Each dot corresponds to one year. x-axis : ζ estimated at the year level following equation (1). y-axis: ζ reaggregated at the year level using a weighted average of "country $\times$ year" estimates, following equation (7). Weights reflect the sample share of each country for the given year.

D Theory Appendix

Proof of Proposition 1

A solution for the ODE given in (D) is:

$$K_a = K_0 e^{(\rho+\beta-\delta)a}$$

Introduce the distribution of sources normalized by the total number of sources for a firm of age a: $f_a = \frac{k_a}{K_a}$. Partially differentiating this distribution with respect to a, and denoting $*$ the convolution product of two distributions yields:

$$\begin{aligned}
\frac{\partial f_a(x)}{\partial a} &= \frac{\frac{\partial k_a(x)}{\partial a}K_a - k_a(x)\frac{\partial K_a}{\partial a}}{(K_a)^2} \\
&= \frac{\left[(\rho-\delta)k_a + \beta\frac{k_a * k_a}{K_a}\right]K_a - k_a(\rho+\beta-\delta)K_a}{(K_a)^2} \\
&= \frac{\beta\left[\frac{k_a * k_a}{K_a} - k_a\right]K_a}{(K_a)^2} \\
&= \beta(f_a * f_a - f_a)
\end{aligned}$$

Using the Fourier transform of f_a yields a simple product instead of a convolution product, which yields that f_a converges towards a Laplace distribution when age grows large (Proposition 2 in Chaney 2018).

One can then derive the endogenized conditions allowing to get a constant elasticity of flows with respect to distance. The distribution of innovator sizes is simply derived from the ODE: $K_a = K_0 e^{(\rho+\beta-\delta)a}$. The relation between a firm's size and its age is $e^a = \left(\frac{K_a}{K_0}\right)^{\frac{1}{\rho+\beta-\delta}}$. With a growth rate of the firm population being equal to γ, this means that the fraction of firms having less than K sources writes:

$$F(K) = 1 - \left(\frac{K}{K_0}\right)^{-\frac{\gamma}{\rho+\beta-\delta}}$$

Thus, the distribution of innovator sizes is Pareto, with a shape parameter $\lambda = \frac{\gamma}{\rho+\beta-\delta}$.

The average squared distance at which firms cite others, Δ_a, is the second moment of the normalized density of sources f_a. Following exactly the steps of the demonstration in Chaney (2018), $\Delta_a = \Delta_0 e^{\beta a}$. Plugging the previous expression $e^a = (\frac{K_a}{K_0})^{\frac{1}{\rho+\beta-\delta}}$, this yields:

$$\Delta(K) = \Delta_0 \left(\frac{K}{K_0}\right)^{\frac{\beta}{\rho+\beta-\delta}}$$

Thus, the average squared distance at which firms cite is a power function of their number of sources, of parameter $\mu = \frac{\beta}{\rho+\beta-\delta}$.