University of East Anglia

CENTRE FOR COMPETITION POLICY

Distance (Still) Hampers Diffusion of Innovations

Georg von Graevenitz
Queen Mary University, CCP and CREATe

Stuart J.H. Graham
Georgia Institute of Technology

Amanda F. Myersc
United States Patent & Trademark Office

CCP Working Paper 19-05R
This version: February 2021

This paper explores a newly emerging source of data on the occurrence and spread of innovations drawn from descriptions of goods and services in the US trademark register. Using these data we describe patterns of regional innovation in the United States and study the effect of distance on the early diffusion of innovations. To identify innovations and their locations we tokenize trademark descriptions and identify novel, fast spreading tokens (words). While trademarks appear to capture many innovations missed in patents, among tokens that co-occur in patent documents the diffusion dynamics for patents and trademarks are similar. We also find that fast growing new tokens are frequently new to English, and their use in language co-evolves with the frequency of linked patents and trademarks. Finally, we show that regional incidence of new tokens parallels patterns of inventive activity reflected in patent data. We exploit occurrence of new trademark tokens to re-examine how spatial distance affects the diffusion of innovations in the U.S. economy. Aggregating innovations at the year and census tract level we estimate Poisson models of diffusion intensity between locations, revealing persistent, strong and negative effects of distance on the intensity of diffusion between locations within the US.

Contact Details:
Georg von Graevenitz g.v.graevenitz@qmul.ac.uk

ISSN 1745-9648

Distance (Still) Hampers Diffusion of Innovations*

Georg von Graevenitz[a†]
a Queen Mary University, CCP and CREATe

Stuart J.H. Graham[b]
b Georgia Institute of Technology

Amanda F. Myers[c‡]
c United States Patent & Trademark Office.

February 1, 2021

ABSTRACT

This paper explores a newly emerging source of data on the occurrence and spread of innovations drawn from descriptions of goods and services in the US trademark register. Using these data we describe patterns of regional innovation in the United States and study the effect of distance on the early diffusion of innovations. To identify innovations and their locations we tokenize trademark descriptions and identify novel, fast spreading tokens (words). While trademarks appear to capture many innovations missed in patents, among tokens that co-occur in patent documents the diffusion dynamics for patents and trademarks are similar. We also find that fast growing new tokens are frequently new to English, and their use in language co-evolves with the frequency of linked patents and trademarks. Finally, we show that regional incidence of new tokens parallels patterns of inventive activity reflected in patent data. We exploit occurrence of new trademark tokens to re-examine how spatial distance affects the diffusion of innovations in the U.S. economy. Aggregating innovations at the year and census tract level we estimate Poisson models of diffusion intensity between locations, revealing persistent, strong and negative effects of distance on the intensity of diffusion between locations within the US.

KEYWORDS: Innovation, Diffusion, Rate of Diffusion, Distance, Innovation Index, Trademarks, Patents

JEL Classification: O3, O51, R1, R32

*We would like to thank Robert Burrell, Amy Cotton, Carsten Fink, Brigitte Granville, Dietmar Harhoff, Michal Kazimierczak, Robert Kimble, Georg Licht, David Muls, Alex Oettl, Pietro Panzarasa, Nathan Wajsman, Kenneth Younge and participants at the 2019 Oxford University Innovation Measures workshop and 2021 GEO INNO conference for their comments on earlier versions of this paper. Graham acknowledges funding from the Center for International Business Education and Research at Georgia Tech.

†Corresponding author: g.v.graevenitz@qmul.ac.uk

‡The views expressed are those of the individual authors and do not necessarily reflect official positions of the Office of Chief Economist or the U.S. Patent and Trademark Office.

1 Introduction

The scholarly consensus holds that geographical distance affects not only the intensity of trade and migration patterns, but also the diffusion of ideas, knowledge and innovations (Clark et al., 2018). Following the early recognition that distance impedes knowledge transfer among people and firms (Marshall, 1920), researchers employed patent data to show that the diffusion of codified knowledge lessens with geographic distance (Henderson et al., 1993). Other researchers have produced similar findings by exploring the transfer of ideas codified in both patents and scientific publications (Peri, 2005; Belenzon and Schankerman, 2013; Singh and Marx, 2013; Li, 2014).

This scholarly consensus has not, however, gone unchallenged. Kolko (2000) notes that the reduction in communication costs and improvements in the speed and quality of interactions might lead to the "death of distance," and possibly an end to agglomeration effects, and finds evidence to support the former. Keller and Yeaple (2013) suggest that, given the intangible nature of ideas, distance may have become less of a barrier for knowledge diffusion after the Internet dramatically lowered communication and interaction costs. Recently, Head et al. (2018) provide empirical support using data on interpersonal networks among mathematicians, finding that over time distance has ceased to create friction limiting idea diffusion in mathematics.

Does a "death of distance" in mathematics reflect a more systematic phenomenon across other disciplines and technologies? We add to the array of data sources used to investigate the spatial diffusion of innovation by exploiting information from the U.S. Patent and Trademark Office (USPTO)'s trademark register, exploring this question across a broad range of innovation domains. Trademark data contain information on the diffusion of information about innovations through the lens of words that identify significant new product and service innovations, such as the smartphone, DVD, blog or outsourcing. The data capture the moment at which products or services are introduced into the market, turning inventions into innovations. Our analysis of this rich and broad set of data on the diffusion of innovations suggest that, while the negative effect of distance may have weakened over time, it is not possible to announce its death by any means.

This paper examines the introduction of new words ("tokens")[1] among the 4.5 million words contained in the USPTO trademark register used to describe goods and services during 1980-2012, and their subsequent re-use. We consider those tokens that are in the top decile of re-use frequency over a period of five years, as a proxy for market impact. For convenience, we employ the nomenclature

[1] We tokenize the goods and services declaration attached to all USPTO trademark applications and list their constituent words. These are the tokens we subsequently study. In other parlance we identify the set of all 1grams within the goods and services declarations.

"innovator" for firms that introduce a new token, and "follower" for subsequent re-use of that token [2]. Using street address information available in the USPTO records, we aggregate new-token emergence and diffusion to the year and 2010 census-tract level.[3] This enables analysis of how geographical distance affects the probability and degree of diffusion of trademarked innovations in the United States from 1980-2012.

The paper first provides descriptive results to support the notion that new tokens in the USPTO trademark register identify market introductions that are, in the parlance of the Oslo Manual (OECD and Eurostat, 2018), "new to the world." We present map visualisations to show that such tokens have primarily arisen in densely populated metropolitan areas associated with innovation activity in the previous literature (Peri, 2005; Forman et al., 2016)[4] and that they tend to diffuse primarily to these same areas[5]. The visualisations also show that the 1996-2005 period saw a substantial increase in the establishment of intensive, long-range diffusion links connecting innovators situated in New York, Los Angeles and the San Francisco Bay Area. To further validate trademark tokens as indicators of innovation, we compare the diffusion of inventions and innovations. Diffusion of inventions is measured through patents associated to the innovation through use of the trademark token in patent titles and abstracts. Diffusion of the innovation is measured through use of the trademark tokens in the trademark data. We observe that the diffusion processes of the inventions and innovations follow similar dynamics and are often closely linked. This suggests that trademark tokens are likely to be informative about diffusion of trademarked innovations in general.

To test the effect of distance on innovation diffusion, we construct panels of directed census tract dyads over 32 years. We analyze incidence of first diffusion and the intensity of diffusion between census tracts. Building on a model of innovation diffusion between areas, we estimate Instrumental Variables Poisson models of diffusion intensity at the level of census tract dyads. The models allow for the possible endogeneity of diffusion links and the ability of locations to absorb innovations. Our primary finding is that distance reduces the intensity of innovation diffusion between locations. This effect was strongest during the period around the Dot com boom. First stage regressions suggest that distance ceased to affect the likelihood that innovation diffused between locations that had not previously been linked consistent with Head et al. (2018).

In sum our findings suggest that tokens derived from goods and services declarations in trademark data contain important information about innovation diffusion drawn from trademark data. Our work

[2]We refer to "followers" because these firms re-use tokens new to the world. Followers may introduce further improvements to the product or service that we do not observe or capture here.

[3]A U.S. census tract covers a geographically contiguous area with an average of 4000 inhabitants. Further information on U.S. census tracts is provided in the online glossary of the U.S. census here. We geocode trademark registrants addresses and link these to 2010 census tracts using spatial merge. The census tract database provides Federal Information Processing Standard (FIPS) codes that identify the county and state each census tract is located in.

[4]Patenting intensive regions in the U.S. include New York City, the San Francisco Bay Area, and Greater Los Angeles.

[5]The 2010 census contained 73,057 separate tracts. 4,295 tracts contain addresses of applicants who introduced fast growing new tokens into USPTO's trademark register. 8,659 tracts contained addresses of firms who took up these tokens within the first year of filing. The vast majority of census tracts contain no innovating firms by this measure.

expands on work by Semadeni (2006) who was the first to use trade mark tokens to analyse innovation, by demonstrating how large sets of innovations can be studied using this source of data.

This paper is structured as follows: In the next section we discuss existing measures of innovation and diffusion. Section 3 contains a descriptive analysis of trademark token data. There we compare dynamics of inventions and innovations linked through tokens. Section 4 introduces an empirical model of diffusion and data based on trademark tokens that we analyse with this model. In Section 5 we present results from estimating the model. Section 6 concludes.

2 Measures of Innovation and Diffusion in Previous Work

This section provides a review of the literature on innovation and diffusion and a discussion of how data from the trademark register can complement other sources of data used to study these phenomena.

2.1 Measuring Innovation

At least since Schumpeter (Schumpeter, 1982 (1934)) recognized a distinction between invention and innovation, researchers have sought meaningful measures of this social phenomenon. The Oslo Manual (OECD and Eurostat, 2018), which sets out a standard for collecting and using innovation data, defines an innovation as "a new or improved product or service." This definition captures not only the elements of novelty and change, but also commercialization.

While several measures have been commonly used in empirical analysis, each suffers from downsides. R&D investment data are readily available and widely used, but measure only inputs to the innovation process. Output measures are more proximate to the market and considered more precise, but are drawn from a limited range of sources: Data from the patent registers provide administrative information on technology inventions, while company surveys are used to track innovation processes within and across firms and universities. Since patent data reflect innovation in a limited range of technologies, surveys present a more comprehensive tool in their coverage of technologies and industries, but are expensive to conduct and usually reflect activity in only a small subset of firms.

Recently, new sources of data have emerged to study innovation. Alexopoulos (2011) and Alexopoulos and Cohen (2011, 2019) exploit data on new book titles covering computers and technology. Hippel et al. (2010) survey consumers in the UK and show that a significant proportion engage in developing and modifying of consumer products. Moser (2012) uses historical data to study innovation beginning in the 1800s when neither patents nor trademarks were widely available. These papers have primarily addressed the question of how much innovation there is (was) as well as when and how innovation arises.

The innovation measure we employ in this paper – derived from the US trademark register – are administrative data similar to patent information, but cover a much broader range of industries. For

instance, trademark protection extends to service industries,[6] an economically vibrant area that patent data largely misses. Trademark data also directly reflect a commercialization event, a characteristic the patent data lack. Furthermore, because trademarks are government data, we are able to link registrant addresses to specific geographical locations.

2.2 Trademark Data as a Source of Information on Innovations

Matthew Semadeni (Semadeni, 2006; Semadeni and Anderson, 2010) uses 252 terms describing consulting services[7] to analyse innovation and competitive interactions of professional services firms. He uses goods and services descriptions from USPTO trademarks to identify innovators and followers. This work demonstrates the potential of goods and services descriptions for the analysis of innovation. In this paper we explore whether his curated approach to the analysis of tokens drawn from goods and services descriptions can be expanded to lists of tokens extracted on the basis of an algorithm. Due to the novelty of this approach we discuss the genesis of goods and services lists here as well as other previous work linking trademark data and innovation.

Graham et al. (2013) note that US law requires that each applicant seeking to register a new trademark must clearly and concisely describe – with specificity – the particular goods and services on which it uses (or intends to use) the mark. The USPTO will accept any of the over 37,000 pre-existing identifications in its catalogue, which range from more specific (”Passenger and light truck tires”) to general (”Tires”). For novel products or services, applicants may also compose their own goods and services identifications. Where there is no common commercial name, the applicant is required to describe the product or service and its intended use. For instance, the mark ”My Trazom” was published in 2011 with the following description: ”Providing a web site that gives computer users the ability to upload, exchange and share photos, videos and video logs.” Another, ”Mablogix” published in 2012, was described as: ”Custom manufacturing and custom synthesis of antibodies and genetically engineered DNA expressing antibodies, biological organisms, cells, viruses, pathogens and special purpose cells to the specifications of others for scientific, research, medical, veterinary and laboratory use.” The USPTO generally does not accept terminology that is overly broad or spans multiple goods and services classes. The specificity of description requirement serves to support not only proper classification and better search, but also gives notice to third parties regarding the scope of an applicant’s rights in a mark. To study innovation we propose to tokenize goods and services lists and then identify new tokens that grow comparatively quickly as described in the following section.

Most previous work linking innovation and trademarks has focused on the trademarks themselves. Since Schmoch (2003) suggested service marks could be used as service innovation indicators,

[6]Due to a change in the trademark classification system introduced in 2000 we have excluded some of the service classes (43-45) from our analysis.

[7]The tokens ”consultancy” and ”consultation” are among the most frequently used type C tokens we identify. The typology of tokens is introduced in Appendix B.

economists have increasingly employed trademark data to provide additional insights to innovation (Mendonca et al., 2004; Ceccagnoli et al., 2010; Flikkema et al., 2014; Thoma, 2015; Graham et al., 2018; Sandro Mendonca, 2019). Trademarks are frequently registered close in time to a new product or service being introduced to the market, so are much nearer to launch date than are patents,[8] and more comparable to the publication dates of technical manuals and books. The trademark registers also reflect activity from a much wider set of sectors and firms than U.S. and European patents (Graham et al., 2013; Dinlersoz et al., 2018). This wider coverage stems from the lower cost of filing trademarks and from their primary objective: to protect a brand or logo against imitation, and to protect consumers from fraud. As such, trademarks are used at least as widely in service industries as they are elsewhere. Moreover, because companies selling physical products frequently do not patent (Moser, 2012; Fontana et al., 2013), trademark data can capture innovation missed in patent data.

2.3 Diffusion of Ideas, Inventions and Innovations

Patent data have been widely available for decades and analysis of how technological inventions diffuse has commonly relied on these data, specifically on patent citation patterns. Jaffe (1986) showed that patent citations could be used to capture knowledge spillovers and Trajtenberg (1990) demonstrated that the number of times a patent is cited captures the technological significance of the patented invention. Henderson et al. (1993) then showed that spillovers of innovations decline with distance[9]. Scientific publications (e.g., journal articles) constitute a separate source of data to study knowledge diffusion. Publications, like patents, contain citations that may reflect reliance on prior ideas.[10] Head et al. (2018) use rich data from mathematics to provide evidence that distance has become less of a barrier over time to the spread of ideas in mathematics. The contrast between their results and the literature on diffusion of technical knowledge may be indicative of differences in diffusion processes or may reflect changes in general conditions affecting diffusion due to new technologies.

Measuring knowledge diffusion using information on technology usage is onerous, because links between the innovation and its (ultimate) use must be inferred. Examples include Comin et al. (2008) and Comin and Hobijn (2010), who provide evidence on the diffusion of a wide range of technologies from data on the adoption of 115 technologies spread over 200 years, across many countries. Comin et al. (2012) also employ these data and find that distance negatively affects technology diffusion.

A similar problem of inference about diffusion mechanisms exists in the data we analyse in this paper. We observe that a term describing a new technology or service appears in the trademark register. The register alone does not reveal that the followers have directly observed the innovator and learned from them about the commercial potential of the inventions being brought to market. But it is documented that the trademark register provides important intelligence for rival firms and that

[8] In pharmaceuticals, patent filing often precedes product introduction by 7-10 years (Grabowski and Vernon, 2000).

[9] This is supported, inter alia, by Peri (2005); Belenzon and Schankerman (2013); Singh and Marx (2013); Li (2014).

[10] However, recent research investigates the incidence of so-called negative citations in science (Catalini et al., 2015).

innovators sometimes go to significant lengths to slow down this avenue for diffusion (Fink et al., 2018). This supports our contention that subsequent users of a trademark token are likely to have learned about potential for an innovation from the leader.

While the literature studying the diffusion of ideas and innovations has continued to grow, there are a number of data-related reasons to use care when interpretating these results. Nelson et al. (2014) argue that firms may over- or under-report adoption of innovations. Moreover, the common use of patent data in these studies can introduce bias since not all sectors or firms rely on patents to protect their innovations (Levin et al., 1987; Cohen et al., 2000; Graham et al., 2009). It is also well established that many patents are not associated with product introductions (Nelson, 2009; Hall and Harhoff, 2012), which may lead to an overestimate of innovation diffusion, though not possibly of ideas, when employing patent-based indicators.

In this respect the trademark data we use for our study present three advantages over patent data. The USPTO requires that the scope of a trade mark is restricted to its use in the market. While this requirement is not always met,[11] it increases the strength of the correlation, relative to patenting, between use of a token in a goods and services declaration and the actual introduction of a corresponding product in the market. A second benefit of using trademark data to study innovation derives from the low cost of applying for trademarks, thereby increasing the range of firms employing trademarks and of innovations reflected in the trademark register.

The third benefit of trademark data is that it covers a much broader range of innovations than patent data: innovation is a broad phenomenon, encompassing technological breakthroughs that are embodied in new products and services, but also new forms of cultural expression, new forms of sport, fashion and language itself. This benefit introduces two challenges. The first concerns breadth: patents are limited to specific inventions under the concept of unity of invention. No such requirement exists for trademark tokens associated to new goods or services. As our examples below demonstrate tokens may be extremely broad (e.g. biotechnology, smartphone) or fairly narrow (e.g. eeprom). This is a feature of the data we have to live with; it must be borne in mind whenever trademark tokens are used to study innovation and will affect the interpretation of any findings as we note below.

A second problem resides in the way we identify innovation: novelty of a token in the set of tokens contained on the trademark register. This means we may identify tokens that are potentially just language innovations, i.e. pre-existing lines of goods and services are described using a new set of words. In these cases there is no product market innovation. To address this second problem we make use of data on word frequencies in natural language derived from an internet based mega-

[11] USPTO conducted a Proof of Use pilot in 2012, requiring trademark owners to submit additional evidence of trademark use on the goods or in connection with the services identified in the registration. In just over half of the randomly selected registrations, owners were unable to verify previously claims of use, resulting in either narrowing of protection through deletion of goods and/or services or outright registration cancellation. See the Post Registration Proof of Use Pilot Final Report (accessed 12 August 2019).

corpus, specifically Google ngram data[12]. We classify tokens representing innovations into six groups, depending on whether they have arisen in natural language before they arise in the set of trademark tokens and whether their usage in natural language changes significantly around the time in which they enter the set of trademark tokens. This classification allows us to measure how significant this second problem is likely to be in each group of trademark tokens.

A limitation of other innovation measures noted by Nelson et al. (2014) also affects trademark tokens: when new goods and services are introduced there may be considerable variation in the terminology used to refer to them across firms. It can take time for a commonly accepted name to emerge. Therefore we may identify the date on which such innovations first emerge incorrectly: the trademark register reveals the date at which the most widely adopted name for an innovation emerges[13].

Balancing these limitations trademark token data offer a range of opportunities. Analysis of trademark tokens can in principle be extended as far back as Britain's first trademark register in 1876 (Bently, 2008). A further benefit derives from the administrative nature of the data: the trademark register reflects both arrival and diffusion of innovations in a way that can be cleanly established. The register contains firms originating new tokens, revealing the spatial concentration of innovations, and the temporal and spatial diffusion to "followers" who subsequently use the same innovation. Moreover, because trademark registration is associated with products and services being introduced in the marketplace, we are arguably focusing on economically or commercially important innovations.

3 Trademark Tokens as Measures of Innovation and Diffusion

In this section we discuss insights into U.S. innovation patterns that can be gleaned using trademark tokens. We use previous findings on innovation derived from patent data to establish whether trademark tokens are likely to capture diffusion of technical innovations. The results provide confidence that trademark tokens are likely to describe diffusion of innovations apart from mere technical innovations.

Trademarks are registered for a broad range of goods and services. If we are to go beyond the analysis of specific industries based on curated token lists (Semadeni, 2006) an algorithm is required that will select large numbers of tokens linked to significant innovations. Here we propose and validate such an algorithm: innovation is reflected in words (tokens) that are entirely new to the corpus of words describing goods or services in the US trademark register. These tokens are "new", on the date they first appear in the register. The subset of tokens most frequently re-used in the register within the subsequent five years is selected, where the cutoff is the ninth decile of the distribution of re-use frequencies within each Nice class during 1980-2012. Details are relegated to Appendix A.

[12] Details regarding this source of data are provided in Appendix B.

[13] A more focused study of only a small subset of innovations in our data could address this issue, but our approach here is to include as wide a range of innovations as possible.

Firms introducing the selected tokens are defined to be innovators, firms reusing the token within the first year are defined to be followers for the purpose of studying diffusion.

This algorithm for selection of tokens identifies terms that are clearly linked to important innovations, e.g. webcast, smartphone, but also others that are less obviously innovative, e.g. consultancy or cremation. To address this we use information from Google's ngram data[14] to further classify the selected tokens. We identify those trademark tokens that are both new to language and new to the trademark register as the most likely candidates for significant novelty. A detailed classification of tokens is developed in Appendix B[15]. The appendix contains extensive lists of tokens of all types with descriptive statistics that measure the degree of novelty and the extent of use of each token.

The algorithm we propose has three important variables: first, the period in which a token can emerge as significant (5 years), second the treshold for significance (9th decile) and third the period in which we identify identify followers (one year). We selected the first to allow us to identify those innovations that grew over a substantial number of years, the second to be conservative with regard to significance and the third to limit the possibility that chains of diffusion arise that we cannot follow with this data alone[16].

This algorithmic approach to the selection of significant tokens cannot ever be as precise as a curated list of tokens: some tokens selected will seem unlikely candidates for a list of significant innovations, even if they are used to describe large numbers of trademarks. In this paper significance is measured relative to usage of the token in patents, trademarks and natural language. We anticipate that such usage is likely correlated with economic importance, but this aspect remains to be explored.

3.1 The Regional Distribution of Innovators

Much of what we have learned about the regional distribution of innovation in the U.S. has come from the patent data. Recently, Forman et al. (2016) study where innovations have arisen, identifying an increase in the share of patents originating from the San Francisco Bay area in California, mainly at the expense of the New York City metro area. Much of the shift they identify occurred between 1990 and 2000, yet was not concentrated in ICT technologies: the re-concentration is more general across technologies at least those reflected in patenting. Hannigan et al. (2015) illustrate the local persistence of automotive innovation in Detroit, Michigan, even as the manufacturing of automobiles has declined in the region. Their analysis documents the simultaneous importance of local and long-distance links to the continued vibrancy of this geographic technology cluster.

We explore both locations of innovators and long-distance diffusion links using data on trademark tokens next. To do this addresses of all firms introducing significant new tokens into the trademark

[14]The data are available on the Google ngram webpages.

[15]Empirical results regarding effects of distance on diffusion of innovations reflected in trademark tokens are not affected by this additional filter. We provide results demonstrating this in Appendix D.

[16]The median token in our data diffuses once in the first 364 days.

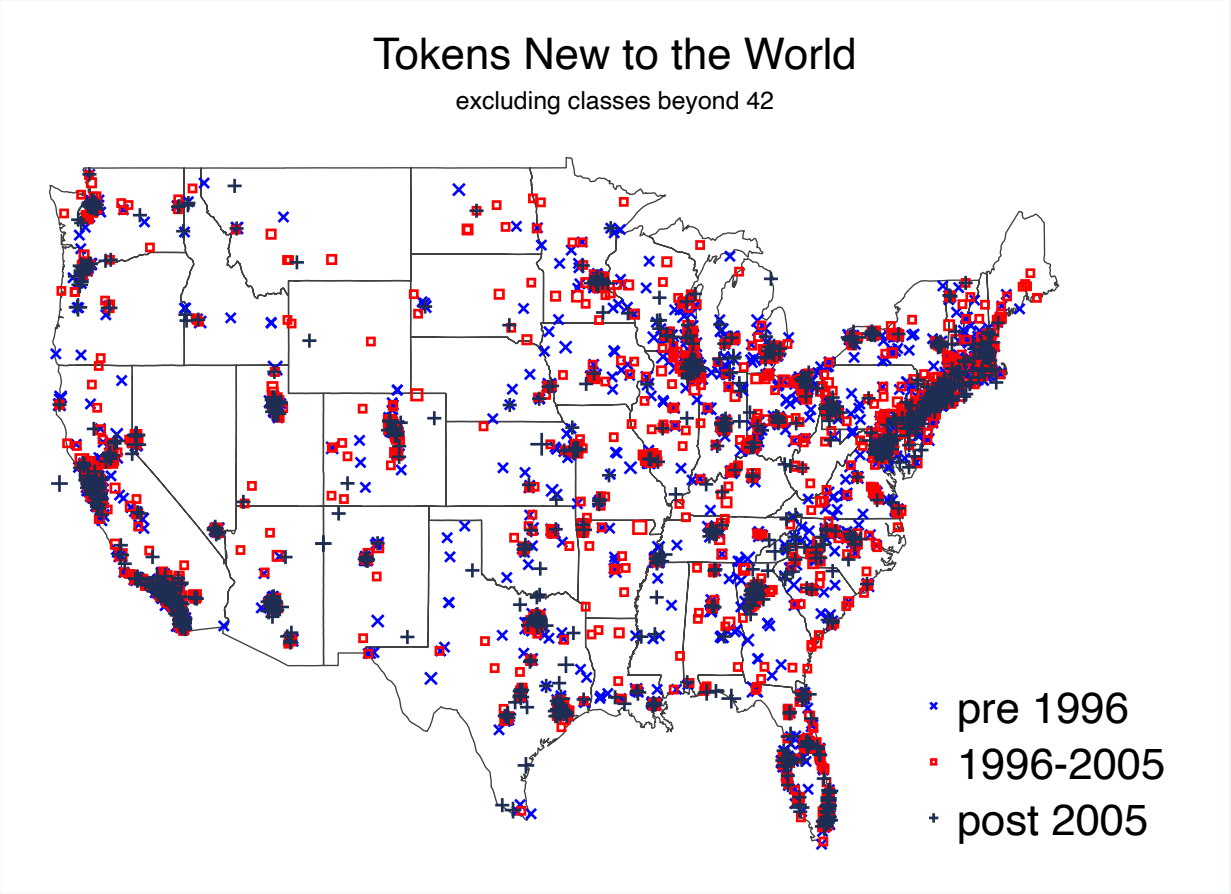

Figure 1

This figure provides a visualization of distinct census tracts contributing significant new word tokens in USPTO trademark descriptions. Symbol size indicates the number of new tokens contributed (larger indicates more tokens) in three time periods (indicated by style and color of symbols). The figure identifies concentrations of innovations in spatial proximity, similar to maps of local labour markets using commuter flows developed by Nelson and Rae (2016). These clusters occur in ten megaregions (e.g. Northeast, Northern California, Southern California) first identified by Lang and Nelson (2007). Due to classification changes, the figure excludes NICE classes 43-45 for consistency.

register (hereafter innovators) are geocoded[17]. Figure 1 shows the distribution of these innovators across the United States, subdivided into three periods to reveal effects of time on the spatial distribution of innovation. We find that 4,295 census tracts out of 73,057 contain innovators. The mapping of innovating census tracts in Figure 1 comports with findings from patent data reported in Forman et al. (2016) and Hannigan et al. (2015): a cluster of innovation in and around the San Francisco Bay Area is visible, as is the cluster in the Northeast megalopolis from Boston to Washington DC and persistent innovation around the rust-belt cities of Minneapolis, MN, Detroit, MI, Cleveland, OH, Pittsburgh, PA, and Buffalo and Rochester, NY.

We extend our analysis by exploiting subsequent uses of fast growing new tokens in the trademark

[17]Geocoding of addresses was done with the help of two Stata modules: `opencagegeo` (Zeigermann, 2016) and `geocodehere` (Hess, 2015). We geocoded a subset of addresses twice and checked the reliability of the packages used.

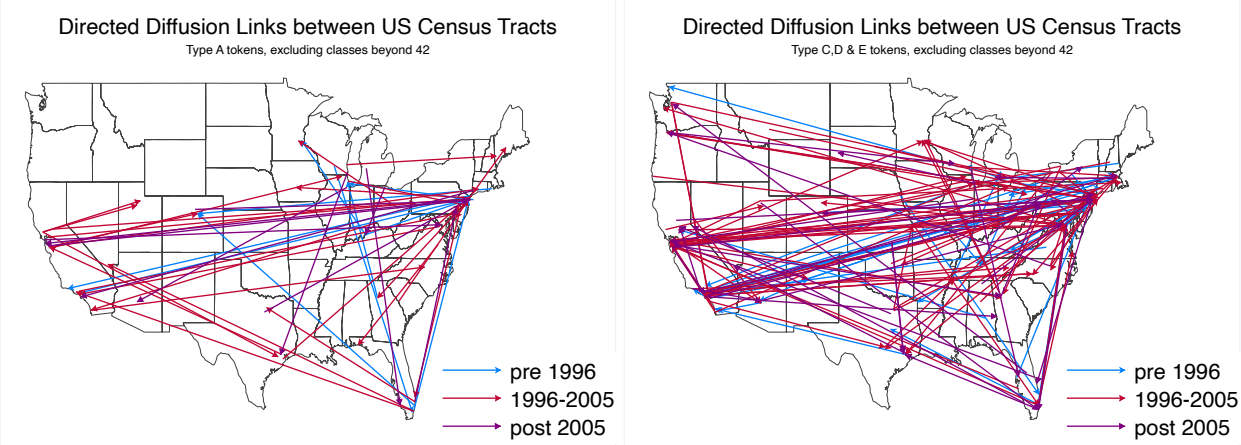

Figure 2

This figure shows arrows starting from addresses at which tokens new to the world are introduced and ending at addresses which subsequently use these tokens within the first year of their introduction. We exclude all distances below 100 miles and only show links across which at least 5 tokens diffused in a given period.

register, and map the diffusion of these tokens within one year of filing. In Appendix B we develop a classification of subtypes of the significant new tokens, exploiting data on usage of words in natural language, specifically books published at different times. We define type-A tokens as those that rarely or never appear in natural language prior to their entry in the trademark register. Types C, D and E appear frequently in books and have varying levels of correlation with a set of frequently used control words pre entry/post entry into the trademark register. Figure 2 shows diffusion links that cover distances of at least 100 miles and across which at least 5 innovations have diffused. Type-A tokens arise in a smaller set of locations and diffuse to fewer locations than innovations linked to tokens of types C, D and E. This shows that these more novel innovations are rarer and arise in even fewer locations than innovations that draw on pre-existing concepts.

Overall Figure 2 illustrates the importance of New York, the San Francisco Bay Area and Greater Los Angeles to the supply of innovation in the United States. In these graphs of diffusion links, the original long-distance diffusion arose primarily between Southern California and New York.

The figures presented here contain more information than we are able to analyze in this article. We note some further avenues for analysis that the figures suggest in the conclusion. Yet it remains to determine whether trademark tokens reveal information about innovations in ways that are closely aligned with more conventionally used innovation measures, such as those derived from patent data. The next section addresses this question.

3.2 Validation Against Patent and Natural Language Measures

This section explores whether the adoption of trademark tokens to describe goods and services follows a similar pattern to those that can be visualized using natural language and patent documents. We expect, consistent with Alexopoulos and Cohen (2011), that as inventions related to a technology become more numerous the technology becomes better known and will be reflected in more references to the technology in books. This characteristic is captured by the ngram data. We also expect that inventions with commercial applications will eventually result in a growing number of trademark applications referencing the token.

We present graphs of cumulative diffusion curves for eight type-A tokens[18], defined as being both new in language and to the trademark register. The graphs contain three diffusion curves for each token: reflecting use in books (natural language), in patents and in trademarks. The tokens presented in Figure 3 are selected from type-A tokens with the highest cumulative patent counts in 2012 (left panel) and with the highest cumulative trademark counts in 2012 (right panel). The selected examples cover digital technology (eeprom, smartphone), nanotechnology (nanoparticle), material sciences (graphene), medicine (prodrugs, immunosuppressants) and services (outsourcing, webcast). In Appendix C we provide four additional cumulative diffusion curves for type-A tokens for which we were unable to find associated patents.

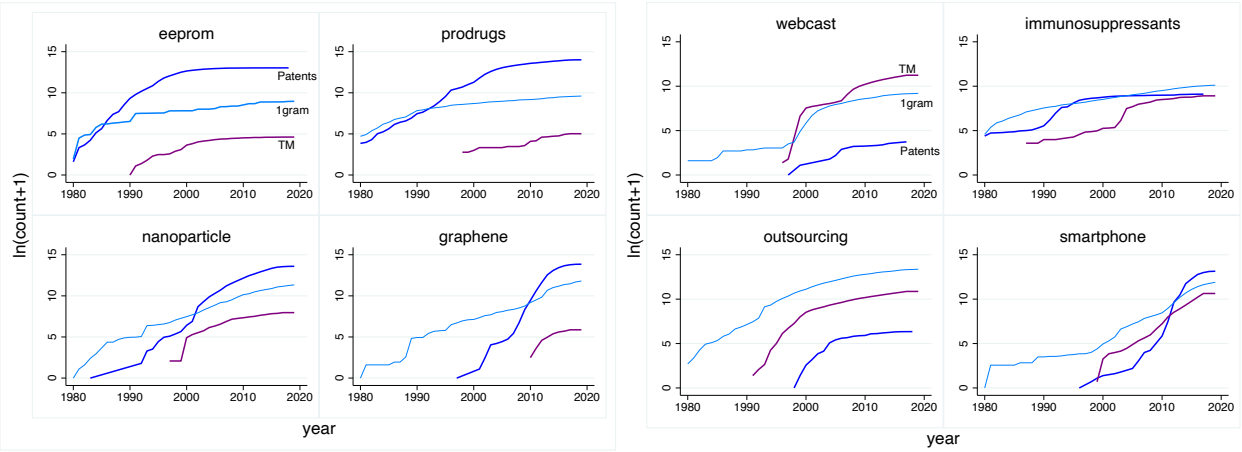

Figure 3: Diffusion of eight innovations captured using token counts from ngram (mid-blue line), patent (dark blue line) and trademark (purple line) data.

The left panel presents data on four innovations associated with high cumulative patent applications, the right panel presents data on four innovations associated with high cumulative trademark applications.

Figure 3 shows that the cumulative diffusion curves for patented inventions and trademarks linked to each token are comparable. The examples offer quite distinct versions of the common s-curve that characterizes diffusion, ranging from the classical (immunosuppressants) to cases in which the take-

[18]There are 2,539 type-A tokens in our data and we link 1,519 (60%) of these to patents.

off phase is missing (eeprom, outsourcing). However, the curves appear correlated with noticeable kinks that co-occur across the data sources (webcast, nanoparticle, eeprom). As should be expected, inventions for the most part precede innovations. Interestingly, exceptions include tokens that appear associated with service innovations (outsourcing). The first use in language (i.e., books) significantly precedes the patenting of inventions for four of the innovations (e.g. graphene), consistent with the idea that published science and academic discussion pre-dates commercial investment.

Overall Figure 3 reveals that diffusion of inventions and innovations are correlated in ways one would expect. This supports the notion that the trademark tokens our algorithm selects reflect innovation. Further empirical study of the diffusion of innovations using trademark token data seems warranted. Since our primary interest here lies with the effects of distance on initial diffusion of innovations, we leave this to future work. The following sections focus on how distance between locations affects the likelihood that innovations diffuse between them.

4 Model and Data

Henderson et al. (1993, 2005) discuss the primary identification problem that affects all studies seeking to estimate how distance affects diffusion of innovation: Is diffusion localized because distance makes it harder to learn about an innovation, or because those most likely to re-use an innovation are located in the same local area as the innovator? In the second case, unobserved factors can generate a cluster that is revealed by diffusion patterns observed in the data. Henderson et al. (1993) address the identification problem by matching patent citations with potential citations that are comparable.[19] Head et al. (2018), who examine personal ties in mathematics, adopt a similar approach to identification as Henderson et al. (1993) to show that the effects of distance on knowledge diffusion in mathematics have decreased over time.

Because the data source analysed here is relatively novel, we do not have sufficiently detailed information on trademark filing entities and their histories to estimate models at the firm level. In a real sense, we suffer from similar teething problems as early researchers working with the patent data experienced. So instead, we aggregate trademark innovations at the census tract level and estimate gravity models of innovation diffusion in census tract dyads. Notably, Peri (2005) and Li (2014) find strong negative effects of distance on knowledge flows using gravity models.

We estimate the capacity of locations to generate and absorb innovations, and borrow from the literature modeling international trade flows to analyze a fixed set of regional links. We find that the number of potential location dyads is far greater than that of innovation-active location dyads.

Becasue the data we analyse span many years, we are able to observe repeated diffusions of innovations between locations over time. Analogous to the way patent-data researchers rely on matching

[19] This approach was subsequently critically tested by Thompson and Fox-Kean (2005) and commented on in Henderson et al. (2005). Singh and Marx (2013) extend the methodology used to estimate this type of matching model.

citing and non-citing locations to construct controls, we include only those census tract pairs for which we observe at least one diffusion event – but include these pairs for the entire 32 years of the sample. We adopt a gravity-model estimation approach suggested by Silva and Tenreyro (2006) allowing us to retain observations for which diffusion counts are zero. We augment this approach by allowing for endogeneity of first diffusion from one location to another, employing lagged variables to instrument both the formation of diffusion links between census tracts and the absorptive capacity of firms located in the receiving census tract.[20]

A further motivation for estimating gravity-type models at the census tract level can be found in the literature on regional innovation systems and clusters. Hannigan et al. (2015) suggest that organizations' capabilities which spawn innovation reside in local and global linkages between clusters. These capabilities endure, and can survive the death or migration of specific entities like firms or research centers which may help to explain our showing of the persistence of innovation clusters in older rust-belt cities like Detroit and Buffalo.

Location specific fixed effects in non-linear panel data models cannot be consistently estimated due to the incidental parameters problem (Lancaster, 2000). We introduce pre-sample data on the number of innovations new to the world arising in the sending and receiving census tracts for the decade 1970-1980 to control for permanent unobservable differences across locations (Hausmann et al., 1984; Blundell et al., 1995).

4.1 Model

We adopt a model of innovation diffusion comparable in spirit to the model proposed by Peri (2005). Diffusion $\mathtt{d}_{s,r,t}$ from a sending ($\boldsymbol{s}$) to a receiving ($\boldsymbol{r}$) area in year $\boldsymbol{t}$ is a function of distance between these areas $\mathtt{D}_{s,r}$, innovation $\mathtt{I}_{s,t}$ in the sending location and absorptive capacity $\mathtt{A}_{r,t}$ in the receiving location and time and area fixed effects $X_{s,r,t}$:

$$\mathtt{d}_{s,r,t} = X_{s,r,t} \left(\mathtt{D}_{s,r}\right)^{\beta_D} \left(\mathtt{I}_{s,t}\right)^{\beta_I} \left(\mathtt{A}_{r,t}\right)^{\beta_A} \quad . \tag{1}$$

This specification can be estimated with a Poisson model. To address the identification problem with which the micro-level literature has grappled (Henderson et al., 1993), we endogenise the first instance of diffusion from sending to receiving location and control for the age of the link between them. We also allow for the endogeneity introduced by a receiving areass absorptive capacity, but assume that the arrival of "new to the world" innovations at the sending location is uncorrelated with location specific unobserved effects[21].

[20] Methods used to endogenise variables in a Poisson framework are developed by Windmeijer and Santos Silva (1997).

[21] We control for location specific constant unobserved effects using pre-sample data on innovation (Blundell et al., 1995).

4.2 Data

We derive the following structural equation from this model:

$$\delta_{s,r,t} = \exp\left(\beta_0 + \beta_D \ln D_{s,r,t} + \beta_I \ln I_{s,t} + \beta_A \ln A_{r,t} + \gamma_L L_{s,r,t} + F'_{s,r,t}\lambda\right) + u_{s,r,t}, \qquad (2)$$

where $\delta_{s,r,t}$ is a count of the number of tokens introduced as new to the world in year $\boldsymbol{t}$ in the **s**ending location which are used within one year at the **r**eceiving location. This is a measure of innovation diffusion between locations: $\mathtt{d}_{s,r,t}$. To endogenise the formation of diffusion links between locations $L_{s,r,t} = 1$ identifies the first year of diffusion in a dyad. When estimating this model we control for time and area fixed effects $F_{s,r,t}$. This equation can be estimated using an Instrumental Variables Poisson model with endogenous covariates and an additive error term (Windmeijer and Santos Silva, 1997; Silva and Tenreyro, 2006).

Equation (2) contains three principal explanatory variables:

Distance$_{s,r,t}$ in miles is calculated as the median distance of all sending - receiving firm pairs per year for each census tract dyad. Distance can vary within a dyad over time to reflect changes in the concentration of economic activity in locations within a census tract. A distance is the geodesic distance between the sending and receiving locations, calculated using the haversine formula. Previous literature suggests that this variable will have a negative effect on diffusion.

Innovation$_{s,t}$ is the count of all new to the world tokens generated in the sending census tract per year. We also use the lagged count of innovation per year in the receiving census tract as an instrument. We do not endogenise innovation.

Net diffusion$_{r,t,}$ is the count of all new to the world tokens diffusing to a receiving census tract, net of those from a sending census tract per year. This variable can be thought of as an analogue of local absorptive capacity for the receiving census code (Cohen and Levinthal, 1990). We instrument this variable.

The median distance between sending and receiving locations in the data is 750 miles. The distribution of distances is U-shaped with a significant spike for very short distances.

In addition, we include several covariates:

Years linked$_{s,r,t}$ measures the number of years that have passed since the first diffusion of an innovation from the sending to the receiving census tract.

Period has three phases: before 1996, between 1996 and 2005 and after 2005. The phases are chosen so as to separate out the decade centered on the Dot com boom in 2000, during which significant investments in Internet mediated communication took place and U.S. trademark registrations were unusually high (Graham et al., 2013).

Table 1: **Descriptive Statistics - type-A & Type B Tokens** (N $= 136,663$)

Variable	**mean**	**sd**	**90th percentile**	**min**	**max**
Diffusion$_{s,r,t}$	0.064	0.944	0	0	231
ln Net Diffusion$_{r,t}$	0.009	0.117	0	0	4.511
Link dummy	0.031	–	0	0	1
ln Years linked$_{s,r,t}$	0.789	1.097	2.833	0	3.466
ln Distance$_{s,r,t}$	5.299	2.812	7.830	0	8.541
ln Innovation$_{s,t}$	0.031	0.161	0	0	4.290
ln Innovation$_{s,t-1}$	0.031	0.160	0	0	4.290
ln Innovation$_{r,t-1}$	0.007	0.080	0	0	4.290
ln Net Diffusion$_{s,t-1}$	0.013	0.176	0	0	5.170
Pre-sample Innovation$_s$	0.883	3.086	6	0	26
Pre-sample Innovation$_r$	0.716	2.283	5	0	26

The panel for type-A&B tokens consists of 4,721 dyads of sending/receiving census tracts. The panel covers the years 1981 to 2012 inclusive.

The algorithm we describe in Section 2.3 identifies 13,749 significant new tokens. 9,227 of the significant new tokens are introduced by a U.S. firm and fall into NICE classes below 43: these are our main sample. Appendix B sets out how tokens can be further subdivided into those that are novel in the English language and those that represent changed uses of language. We identify 3,645 tokens that occur infrequently in the English language prior to introduction of the token into the USPTO trademark database. Of these 2,386 tokens remain in the data once we restrict to classes below 43 and US applicants. Appendix B contains extensive lists of these type-A and type B tokens as well as of the remaining categories of tokens we identify. The analysis presented below is based on the 2,386 type-A type B tokens. Appendix C replicates results based on all 9,227 tokens.

The 2,386 type-A & B tokens originate from addresses in 1,127 distinct census tracts[22] and are received by (diffuse to) addresses in 2,533 distinct census tracts. For 15.6% of sending-receiving census tract dyads, the sending and receiving addresses lie within the same census tract.[23]

Balanced panels of census tract and year aggregates of new to the world token introductions and their diffusion are used in our analysis below. We identify 4,271 census tract dyads having at least

[22]Data on census tract boundaries was obtained from ESRI. It can be downloaded here. We used the Stata add-on `geoinpoly` Picard (2015) for the spatial merge of firm locations to census tracts.

[23]This does not affect our analysis of distance effects as we construct the distance between sending and receiving addresses per census tract dyad as the median distance within each dyad and year. Having geocoded the addresses of all filing entities we are able to construct distances even within the same census tract.

one instance of a diffusion from sending to receiving census tract in the 32 years between 1981 and 2012. Table 1 sets out descriptive statistics for the resulting panel.

5 Results

This section contains estimation results for the sample of type-A and type-B tokens. We begin with results from the first stage models and then present and discuss results from estimation of the structural equation. Appendix, Section D, contains descriptive statistics and results for the full sample of all tokens identified by the algorithm described in Section 2.3.

5.1 First Stage Models: Probability of Diffusion and Net Diffusion

When estimating the structural equation (2) we allow for the endogeneity of link formation between dyads of sending and receiving census tracts. We also allow for the endogeneity of the receiving location's absorption of innovations from other census tracts. Here we discuss results from estimation of first stage models for first diffusion and net diffusion of the receiving census tract in Table 2.

We condition on the distance between sending and receiving census tracts as well as the innovation rate in the sending census tract. Table 2 sets out two versions of each model: The first contains no interactions, while the second allows for interactions between distance and time periods. We introduce four variables to instrument the endogenous variables: lagged net diffusion to the receiving census tract, lagged net diffusion to the sending census tract, lagged innovation in the receiving census tract and age of the link between census tracts. We also condition on pre-sample innovation levels in both census tracts.

The probability of link formation between census tracts decreases with the median distance between all sending and receiving firms in the two locales. While this effect is statistically significant, it is very small in absolute value: a one standard deviation increase of $\ln$ distance from the mean (53%) would decrease the probability of first diffusion by six hundredths of a percent. The mean probability of first diffusion in the sample is 50 times higher than this (Model 2). Moreover, the interaction with dummies for periods two (1996-2005) and three (post 2005) results in insignificant effects for distance in those two periods (Model 2). These results support the notion that distance is not, or is no-longer, an meaningful source of friction for creation of diffusion links between even fairly remote areas in the United States. This finding is notable given the countrys large land area (9.8 million km^2).

Table 2 also shows that innovation in the sending census tract increases the probability that innovations will diffuse between tracts. Interestingly however, a higher rate of diffusion in the past from other census tracts to the receiving census tract reduces that probability of diffusion, suggesting that persistent transfer links between specific locales may pose barriers to entry to ideas and innovations from different locations. Effects in the first stage models for net diffusion to the receiving census tract

are largely similar to effects just discussed.

Table 2: First Stage Models: First Diffusion and Net Diffusion (N=136,663)

Dependent Variable	First Diffusion$_{s,r}$		Net Diffusion$_r$	
Distance & Periods	in levels	interacted	in levels	interacted
	(1)	(2)	(3)	(4)
ln Distance$_{s,r,t}$	-0.0003^{***}	-0.0012^{***}	-0.0002	-0.0005^{***}
	(0.0001)	(0.0002)	(0.0001)	(0.0001)
ln Innovation$_{s,t}$	0.3254^{***}	0.3256^{***}	0.1039^{***}	0.1040^{***}
	(0.0160)	(0.0160)	(0.0112)	(0.0112)
ln Innovation$_{s,t-1}$	0.3203^{***}	0.3205^{***}	0.0769^{***}	0.0769^{***}
	(0.0137)	(0.0137)	(0.0069)	(0.0069)
ln Net Diffusion$_{r,t-1}$	-0.0332^{***}	-0.0333^{***}		
	(0.0071)	(0.0071)		
ln Net Diffusion$_{s,t-1}$	-0.0961^{***}	-0.0962^{***}	-0.0243^{***}	-0.0243^{***}
	(0.0055)	(0.0055)	(0.0030)	(0.0030)
ln Innovation$_{r,t-1}$			-0.0193^{*}	-0.0190
			(0.0098)	(0.0098)
ln Age of link$_t$			-0.0075^{***}	-0.0075^{***}
			(0.0006)	(0.0006)
Pre-sample Innovation$_s$	-0.0023^{***}	-0.0023^{***}	-0.0005^{***}	-0.0005^{***}
	(0.0002)	(0.0002)	(0.0001)	(0.0001)
Pre-sample Innovation$_r$	0.0001	0.0001	0.0022^{***}	0.0022^{***}
	(0.0001)	(0.0001)	(0.0002)	(0.0002)
Period 2 $\times$ ln Distance$_{s,r,t}$		0.0013^{***}		0.0004
		(0.0004)		(0.0003)
Period 3 $\times$ ln Distance$_{s,r,t}$		0.0021^{***}		0.0006^{*}
		(0.0003)		(0.0002)
Constant	0.0156^{***}	0.0156^{***}	0.0097^{***}	0.0096^{***}
	(0.0007)	(0.0007)	(0.0008)	(0.0008)
R^2	0.1851	0.1852	0.0446	0.0447

[1] Robust standard errors clustered at county dyad level in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[2] All models include time fixed effects.

5.2 Instrumental Variables Models for Diffusion of Innovations

Table 3: IV Poisson Models of Innovation Diffusion - Sample: type-A & B Tokens (N=136,663)

Model	no IV	Restricted IV		Full IV	
Distance & Periods		in levels	interacted	in levels	interacted
First Diffusion (1/0)	5.1908^{***}	7.3836	7.5669	4.8741^{***}	5.4256^{***}
	(0.1596)	(3.8589)	(4.4523)	(1.1183)	(0.9108)
$\ln$ Net Diffusion$_{r,t}$	0.3745^{***}	1.0893^{*}	1.1234^{*}	3.1076^{***}	2.6137^{**}
	(0.0644)	(0.5488)	(0.5731)	(0.9443)	(0.7957)
$\ln$ Distance$_{s,r,t}$	-0.1126^{***}	-0.1070^{***}	-0.1302^{***}	-0.2277^{***}	-0.2285^{**}
	(0.0174)	(0.0202)	(0.0292)	(0.0679)	(0.0789)
$\ln$ Innovation$_{s,t}$	0.8374^{***}	0.2946	0.2749	-0.8402	-0.6203
	(0.1204)	(0.3836)	(0.4004)	(0.4483)	(0.4271)
Pre-sample Innovation$_{s}$	-0.0012	0.0225	0.0221	0.0816^{*}	0.0709^{*}
	(0.0074)	(0.0139)	(0.0143)	(0.0372)	(0.0335)
Pre-sample Innovation$_{r}$	0.0201^{*}	-0.0433	-0.0465	-0.2512^{**}	-0.2075^{*}
	(0.0096)	(0.0549)	(0.0583)	(0.0973)	(0.0882)
Constant	-4.9059^{***}	-6.3908	-6.4812	-3.7530^{***}	-4.1160^{***}
	(0.2602)	(3.8059)	(4.4041)	(0.6887)	(0.8606)
			Marginal Effects		
$\ln$ Distance$_{\text{pre 1996}}$	-0.0037^{***}		-0.0043^{***}		-0.0070^{**}
	(0.0008)		(0.0010)		(0.0026)
$\ln$ Distance$_{1996-2005}$	-0.0099^{***}		-0.0117^{***}		-0.0203^{*}
	(0.0016)		(0.0027)		(0.0085)
$\ln$ Distance$_{\text{post 2005}}$	-0.0038^{**}		-0.0059^{*}		-0.0124
	(0.0014)		(0.0024)		(0.0068)

[1] Robust standard errors clustered at county dyad level in parentheses: * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[2] All models include time fixed effects.

[3] Common instruments: $\ln$ Innovation$_{r,t-1}$, $\ln$ Net Diffusion$_{r,t-1}$, $\ln$ Net Diffusion$_{s,t-1}$

Table 3 sets out results from estimating the structural equation for diffusion intensity between locations: Equation 2. There are five columns in the table: the first provides a baseline model in which we use no instruments. The restricted IV results are obtained by excluding age of the link between

locations from the set of instruments. This variable has a significant negative effect on net diffusion to the receiving location (see Table 2). Table 3 also provides marginal effects for the distance variable, by time period, below the main results.

Distance has a statistically significant large and negative effect on the number of innovations that diffuse from a sending to a receiving location: a five percent increase in the distance between locations reduces diffusion by over 1%.

Net diffusion to a receiving location increases the number of innovations generated in the sending location that diffuse to the receiving census tract (suggesting an absorptive capacity pull effect). This coefficient is severely downward biased when we do not instrument first diffusion and net diffusion to the receiving census tract.

Controls for pre-sample innovation levels in both sending and receiving locations significantly affect diffusion. This indicates the importance of unobserved location specific effects for diffusion.

Allowing for the endogeneity in the structural equation affects the size of the marginal effects we estimate for distance. However, the pattern of effects is the same: distance impeded diffusion almost three times as much during the years of the Dot com boom for these type-A and type-B innovations. We find the same pattern holds when we analyze diffusion for the full set of trademark tokens identified by the algorithm introduced in Section 3. These findings are relegated to Appendix D. This suggests that the greater dynamism of that period produced innovations that were less transferable to businesses at greater distance from the originators.

To summarize, our results show that spatial distance no longer affects the creation of diffusion links after 1996. However, contingent on previous diffusion from a sending to a receiving census tract, we find persistent, strong and negative effects of greater spatial distance on the intensity (extent) of diffusion for existing transfer links between locations.

6 Conclusion

This paper contains new evidence on the effect that geographical distance has on the diffusion of innovations. A primary contribution is the description and application of a previously unused source of information on innovations and their diffusion, namely the emergence and re-use of ”new to the world” terms (tokens) contained in the goods and services descriptions of administrative trademark registrations. While this paper considers trademark information generated during only three decades in the United States, there is wide scope for this measure to be constructed from public trademark information in any country, and for periods beginning as early as the late-1800s, when administrative trademark registers began to be recorded.

While the consensus scholarly view is that the diffusion of ideas and innovation decreases as spatial distance increases, recent scholarship questions this regularity. By linking new trademark tokens to the business addresses of innovator and follower firms, and defining substantial innovations

as new to the world tokens in goods and services descriptions, we are able to analyze the diffusion patterns of the most commercially impactful innovations linked to trademarks from 1980 through 2012.

Our results largely confirm findings in previous work, which has shown that distance hampers the spread of ideas, inventions and innovations. The novelty of our findings lies in the source of information about innovation and diffusion, which is entirely different from that exploited in previous work. We address some of the endogeneity likely to affect this type of work. However, we do not have experimental or quasi-experimental data at our disposal and are not in a position to control for other pathways, such as the personal networks which may mediate the diffusion of innovations. Therefore we cannot entirely rule out that diffusion is primarily local, because even in the age of the Internet, the networks of innovators and gatekeepers in the sense of Roberts and Fusfeld (1982) may remain primarily local.

The data source we exploit captures innovation for a much broader range of applications and technologies than those reflected in patent data, most innovation surveys, and even the books and manuals recently studied by Alexopoulos and Cohen (2011). However, this breadth also introduces certain limitations: the breadth of the data means that diffusion mechanisms that explain how information passes between firms may vary. The breadth of these data derives from its administrative nature, but this also means that we have reliable information only about the location of the firm that handles the registration of trademarks. These data can teach at best a limited amount concerning the process from invention to innovation, and is silent about the location of the inventors who contributed to the innovation. However, these limitation may be remedied in further research by studying in greater detail the links between patents and trademarks, matched through common tokens and filed by the same firm. A further dimension of the data we have not studied here is whether significant references to a token in multiple sources (for instance language, patents and trademarks) are indicative of economic impact, as measured using product sales, employment growth or the like. Finally, the algorithm we adopt excludes interesting innovation phenomena, such as sleeping beauties (Ke et al., 2015): inventions that do not find widespread commercial use or scientific recognition for a long period of time. Future researchers could remedy this shortcoming by carefully integrating data on scientific publications, patents and trademarks based on common tokens, over longer periods of time than we use here.

Beyond these questions related to diffusion, trademark tokens open opportunities for new research into aggregate innovation activities in an economy. The trademark token data are particularly useful for analysis of technologies, industries, firms, and economies for which patents – the most common data used to analyze diffusion – are less suitable. Accordingly, our descriptive analysis of the trademark token data suggests that further work is warranted to better understand what additional insights trademark tokens can reveal about diffusion patterns post-1876. If it can be established that the propensity for new tokens describing new technologies has remained relatively constant over the period after 1876, trademark tokens would be useful indicators, adding insight into the amount of

innovation generated since the late 1800s. Such a finding would help bring light to many important questions, including possibly those being raised in the recent literature about the productivity slowdown affecting advanced economies (Gordon, 2018).

Future analysis would also be welcome as regards the descriptive results presented in the paper, suggesting that the locations of innovating firms have become more concentrated over time. Careful analysis of the number of new firms and their locations is warranted, possibly using other data sources to add precision and depth. Finally, the analysis we have provided here neglects the question of technology fields and clusters of inventive activity, as well as clusters of firms innovating in similar product markets. Again, future research would require a more accurate classification of economic activities than presented by the Nice classification, which are themselves a relatively blunt tool. As matching patent classifications to industry codes vexed early researchers using that data source, trademark applications and their associated product introductions could benefit from more accurate classification, thus presenting research opportunities for future investigation.

References

ALEXOPOULOS, M. (2011): "Read All about It!! What Happens Following a Technology Shock?" *The American Economic Review*, 101, 1144–1179.

ALEXOPOULOS, M. AND J. COHEN (2011): "Volumes of Evidence: Examining Technical Change in the Last Century through a new Lens," *Canadian Journal of Economics/Revue canadienne d'économique*, 44, 413–450.

——— (2019): "Will the New Technologies Turn the Page on US Productivity Growth?" *Economics Letters*, 175, 19–23.

BELENZON, S. AND M. SCHANKERMAN (2013): "Spreading the Word: Geography, Policy, and Knowledge Spillovers," *Review of Economics and Statistics*, 95, 884–903.

BENTLY, L. (2008): "The Making of Modern Trade Mark Law: The Construction of the Legal Concept of Trade Mark (1860–1880)," in *Trade Marks and Brands: An Interdisciplinary Critique*, ed. by L. Bently, J. Davis, and J. C. Ginsburg, Cambridge: Cambridge University Press, chap. 1, 3–41.

BLUNDELL, R., R. GRIFFITH, AND J. V. REENEN (1995): "Dynamic count data models of technological innovation," *The Economic Journal*, 333—344.

CATALINI, C., N. LACETERA, AND A. OETTL (2015): "The incidence and role of negative citations in science," *Proceedings of the National Academy of Sciences*, 112, 13823–13826.

CECCAGNOLI, M., S. J. GRAHAM, M. J. HIGGINS, AND J. LEE (2010): "Productivity and the role of complementary assets in firms demand for technology innovations," *Industrial and corporate change*, 19, 839–869.

CLARK, G. L., M. P. FELDMAN, M. S. GERTLER, AND D. WÓJCIK (2018): *The New Oxford Handbook of Economic Geography*, Oxford University Press.

COHEN, W. M. AND D. A. LEVINTHAL (1990): "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative science quarterly*, 35, 128–152.

COHEN, W. M., R. R. NELSON, AND J. P. WALSH (2000): "Protecting their Intellectual Assets: Appropriability Conditions and Why U.S. Manufacturing Firms Patent (or Not)," Working Paper 7552, NBER.

COMIN, D. AND B. HOBIJN (2010): "An Exploration of Technology Diffusion," *American Economic Review*, 100, 2031–59.

COMIN, D., B. HOBIJN, AND E. ROVITO (2008): "A new approach to measuring technology with an application to the shape of the diffusion curves," *The Journal of Technology Transfer*, 33, 187–207.

COMIN, D. A., M. DMITRIEV, AND E. ROSSI-HANSBERG (2012): "The Spatial Diffusion of Technology," Tech. rep., National Bureau of Economic Research.

DINLERSOZ, E. M., N. GOLDSCHLAG, A. MYERS, N. ZOLAS, ET AL. (2018): "An Anatomy of US Firms Seeking Trademark Registration," Tech. rep.

FINK, C., A. FOSFURI, C. HELMERS, AND A. F. MYERS (2018): *Submarine Trademarks*, vol. Economic Research Working Paper No. 51, WIPO.

FLIKKEMA, M., A.-P. DE MAN, AND C. CASTALDI (2014): "Are trademark counts a valid indicator of innovation? Results of an in-depth study of new benelux trademarks filed by SMEs," *Industry and Innovation*, 21, 310–331.

FONTANA, R., A. NUVOLARI, H. SHIMIZU, AND A. VEZZULLI (2013): "Reassessing patent propensity: Evidence from a dataset of R&D awards, 1977–2004," *Research Policy*, 42, 1780–1792.

FORMAN, C., A. GOLDFARB, AND S. GREENSTEIN (2016): "Agglomeration of Invention in the Bay Area: Not just ICT," *American Economic Review*, 106, 146–51.

FRIGINAL, E., M. WALKER, AND J. B. RANDALL (2014): "Exploring Mega Corpora: Google Ngram Viewer and the Corpus of Historical American English," *EuroAmerican Journal of Applied Linguistics and Languages*, 1, 48–68.

GORDON, R. J. (2018): "Declining American economic growth despite ongoing innovation," *Explorations in Economic History*, 69, 1–12.

GRABOWSKI, H. G. AND J. M. VERNON (2000): "Effective Patent Life in Pharmaceuticals," *International Journal of Technology Management*, 19, 98–120.

GRAHAM, S. J., G. HANCOCK, A. C. MARCO, AND A. MYERS (2013): "The USPTO Trademark Case Files Dataset: Descriptions, Lessons, and Insights," *USPTO Working Paper (January 31)*.

GRAHAM, S. J., R. P. MERGES, P. SAMUELSON, AND T. SICHELMAN (2009): "High technology entrepreneurs and the patent system: Results of the 2008 Berkeley patent survey," *Berkeley Technology Law Journal*, 24, 1255.

GRAHAM, S. J. H., A. C. MARCO, AND A. F. MYERS (2018): "Monetizing Marks: Insights from the USPTO Trademark Assignment Dataset," *Journal of Economics & Management Strategy*, 27, 403–432.

HALL, B. H. AND D. HARHOFF (2012): "Recent Research on the Economics of Patents," Working Paper 17773, National Bureau of Economic Research.

HANNIGAN, T. J., M. CANO-KOLLMANN, AND R. MUDAMBI (2015): "Thriving Innovation Amidst Manufacturing Decline: The Detroit Auto Cluster and the Resilience of Local Knowledge Production," *Industrial and Corporate Change*, 24, 613–634.

HAUSMANN, J., B. HALL, AND Z. GRILICHES (1984): "Econometric Models for Count Data with an Application to the Patents- R&D Relationship," *Econometrica*, 52, 909–938.

HEAD, K., Y. A. LI, AND A. MINONDO (2018): "Geography, Ties, and Knowledge Flows: Evidence from Citations in Mathematics," *Review of Economics and Statistics*.

HENDERSON, R., A. JAFFE, AND M. TRAJTENBERG (2005): "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment: Comment," *American Economic Review*, 95, 461–464.

HENDERSON, R., A. B. JAFFE, AND M. TRAJTENBERG (1993): "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations," *Quarterly Journal of Economics*, 108, 577–598.

HESS, S. (2015): "GEOCODEHERE: Stata module to provide geocoding relying on Nokias Here Maps API," Statistical Software Components, Boston College Department of Economics.

HIPPEL, E. V., J. D. JONG, AND S. FLOWERS (2010): "Comparing Business and Household Sector Innovation in Consumer Products: Findings from a Representative Study in the UK," *Social Science Research Network*, 1–39.

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

KE, Q., E. FERRARA, F. RADICCHI, AND A. FLAMMINI (2015): "Defining and Identifying Sleeping Beauties in Science," *Proceedings of the National Academy of Sciences*, 112, 7426–7431.

KELLER, W. AND S. R. YEAPLE (2013): "The Gravity of Knowledge," *American Economic Review*, 103, 1414–44.

KOLKO, J. (2000): *The Internet Upheaval: Raising Questions, Seeking Answers in Communications Policy*, MIT press Cambridge, MA, chap. 4.: The Death of Cities? The Death of Distance? Evidence from the Geography of Commercial Internet Usage, 73–98.

LANCASTER, T. (2000): "The incidental parameter problem since 1948," *Journal of Econometrics*, 95, 391 – 413.

LANG, R. E. AND A. C. NELSON (2007): "America 2040: The Rise of the Megapolitans," *Planning*, 73, 7–12.

LEVIN, R. C., A. K. KLEVORICK, R. R. NELSON, AND S. G. WINTER (1987): "Appropriating the Returns from Industrial Research and Development," *Brookings Papers on Economic Activity*, 3.

LI, Y. A. (2014): "Borders and Distance in Knowledge Spillovers: Dying over Time or Dying with Age? - Evidence from Patent Citations," *European Economic Review*, 71, 152–172.

MARSHALL, A. (1920): *Principles of Economics*, Macmillan London.

MENDONCA, S., T. S. PEREIRA, AND M. M. GODINHO (2004): "Trademarks as an Indicator of Innovation and Industrial Change," *Research Policy*, 33, 1385 – 1404.

MOSER, P. (2012): "Innovation without Patents: Evidence from World's Fairs," *The Journal of Law and Economics*, 55, 43–74.

NELSON, A. J. (2009): "Measuring Knowledge Spillovers: What Patents, Licenses and Publications Reveal about Innovation Diffusion," *Research policy*, 38, 994–1005.

NELSON, A. J., A. EARLE, J. HOWARD-GRENVILLE, J. HAACK, AND D. YOUNG (2014): "Do Innovation Measures Actually Measure Innovation? Obliteration, Symbolic Adoption, and Other Finicky Challenges in Tracking Innovation Diffusion," *Research Policy*, 43, 927–940.

NELSON, G. D. AND A. RAE (2016): "An Economic Geography of the United States: From Commutes to Megaregions," *PloS One*, 11, e0166083.

OECD AND EUROSTAT (2018): *Oslo Manual 2018*.

PERI, G. (2005): "Determinants of Knowledge Flows and their Effect on Innovation," *Review of Economics and Statistics*, 87, 308–322.

PICARD, R. (2015): "GEOINPOLY: Stata Module to Match Geographic Locations to Shapefile Polygons," Statistical Software Components, Boston College Department of Economics.

ROBERTS, E. B. AND A. R. FUSFELD (1982): "Critical functions: Needed Roles in the Innovation Process," in *Career Issues in Human Resource Management*, ed. by R. Katz, Prentice-Hall, Englewood Cliffs, NJ.

SANDRO MENDONCA, ULRICH SCHMOCH, P. N. (2019): *Springer Handbook of Science and Technology Indicators*, Springer Nature, chap. Interplay of Patents and Trademarks as Tools in Economic Competition.

SCHMOCH, U. (2003): "Service Marks as Novel Innovation Indicator," *Research Evaluation*, 12, 149–156.

SCHUMPETER, J. A. (1982 (1934)): Edison, NJ: Transaction Books.

SEMADENI, M. (2006): "Minding your Distance: How Management Consulting Firms use Service Marks to Position Competitively," *Strategic Management Journal*, 27, 169–187.

SEMADENI, M. AND B. S. ANDERSON (2010): "The Follower's Dilemma: Innovation and Imitation in the Professional Services Industry," *Academy of Management Journal*, 53, 1175–1193.

SILVA, J. S. AND S. TENREYRO (2006): "The Log of Gravity," *The Review of Economics and statistics*, 88, 641–658.

SINGH, J. AND M. MARX (2013): "Geographic Constraints on Knowledge Spillovers: Political Borders vs. Spatial Proximity," *Management Science*, 59, 2056–2078.

THOMA, G. (2015): "The Value of Patent and Trademark Pairs," in *Academy of Management Proceedings*, Academy of Management Briarcliff Manor, NY 10510, vol. 2015, 12373.

THOMPSON, P. AND M. FOX-KEAN (2005): "Patent Citations and the Geography of Knowledge Spillovers: A Reassessment," *American Economic Review*, 95, 450–460.

TRAJTENBERG, M. (1990): "A Penny for your Quotes: Patent Citations and the Value of Innovations," *Rand Journal of Economics*, 21, 172–187.

WINDMEIJER, F. A. AND J. M. SANTOS SILVA (1997): "Endogeneity in Count Data Models: An Application to Demand for Health Care," *Journal of Applied Econometrics*, 12, 281–294.

YOUNES, N. AND U.-D. REIPS (2019): "Guideline for Improving the Reliability of Google Ngram Studies: Evidence from Religious Terms," *PloS one*, 14, e0213554.

ZEIGERMANN, L. (2016): "OPENCAGEGEO: Stata Module for Forward and Reverse Geocoding Using the OpenCage Geocoder API," Statistical Software Components, Boston College Department of Economics.

APPENDIX

A Innovation and Diffusion Indicators using Trademark Tokens

All trade mark applications contain a description of the goods and services which the mark will be used for. These descriptions contain widely used terms that are familiar to both the examiners at USPTO and trade mark attorneys. The use of such standard terms simplifies disputes about overlap of goods and services and makes translation easier. We refer to the list of all terms used to describe goods and services as a corpus of tokens and exploit the introduction of new tokens into this corpus[24].

Goods and services descriptors attached to each trade mark application filed at USPTO provide a good approximation to the range of products marketed under the trade mark name. This is due to the requirement that applicants provide USPTO with proof of the use of their marks in commerce as they are described in the application. When applicants introduce entirely new types of products or services, e.g. GPS, they will often also need to introduce new tokens into the corpus. Should the market for such new products and services grow we would expect many firms entering these markets and filing marks to take up new tokens in their own filings. In order to identify innovation we study the introduction of those new tokens in the corpus that are subsequently widely adopted.

From data about such new, fast growing terms we extract the fastest growing 10% applying an algorithm with four steps:

1. determine the year in which new tokens first appear in the corpus of all tokens in a given Nice class;

2. obtain the frequency with which new tokens are used in the first five years within that Nice class;

3. obtain the 9th decile of the new tokens frequency distribution, where this distribution is constructed relative to the Nice class and the sub-corpus of all new tokens in that Nice class across all years;

4. retain those tokens used more frequently than the 9th decile in the distribution of frequencies.

This algorithm contains a number of parameters that can be adjusted, such as the number of years over which the impact of the new tokens is measured and the quantile used to define significant tokens. Exploration of variants is left to future work.

We restrict our sample of diffusion events to subsequent uses of each significant token within the 364 days after its introduction into the corpus. This restriction is adopted to limit the extent to which

[24] In natural language processing a collection of texts used as a basis for a descriptive analysis is referred to as a *corpus* of text. The term *token* is used to refer to individual words within such a corpus.

addresses had to be geocoded. It also has the benefit of limiting the likelihood that diffusion takes place indirectly[25].

B Classification of Trademark Tokens Based on Ngrams

Appendix A sets out the algorithm used to select trademark tokens based solely on data obtained from the trademark register. This section sets out a method to classify the novelty of a trademark token using data on word frequencies in English. The frequency with which a token has been used in natural language before and after the token arises in the trademark corpus provides additional information on the type of innovation that is represented by the token.

To capture use of the token in natural language we use a word frequency database. Due to the level of documentation available[26] and size of the files this paper is based on Google's 1-gram data released for British English in February 2020[27].

Based on analysis of the co-occurrence of tokens in the Google 1-gram and USPTO trademark corpora we distinguish six levels of innovation. These are set out in Table 4:

Table 4: **Innovation Levels Captured by Trademark Tokens**

Innovation Level	Definition	Count	Percent
A	No record of token in 1-gram corpus in at least 20 years after 1940.	2,539	18.47
B	Token present in 1-gram corpus at low† frequency prior to use in trademark corpus; significant† growth of use in 1-gram corpus thereafter.	1,106	8.04
C	Token present in 1-gram corpus, low correlation with controls*.	3,625	26.37
D	Token present in 1-gram corpus, low correlation with controls* after introduction into trademark corpus.	407	2.96
E	Token present in 1-gram corpus, low correlation with controls* before introduction into trademark corpus, then high correlation with controls.	1,216	8.84
X	All remaining tokens.	4,856	35.32
Total		13,749	100.00

[25] In the data we use in the paper and in the data used in Appendix D the median token diffuses once in this period. Tokens at the 95th percentile diffuse 7 times/ 8 times respectively in these data

[26] Alternative ngram data are available on the pages of the COCA project and those of the Hathi Trust. We intend to explore the use of these data in future work. For discussion of the use of mega-corpora to study language, refer to Friginal et al. (2014) and Younes and Reips (2019).

[27] The data are available on the Google ngram webpages.

† low frequency is defined below the 25th percentile of matches relative to all tokens in the data and significant growth is growth above the 75th percentile of all tokens in the data.

⋆ We use a set of frequently occurring control words suggested by Younes and Reips (2019) to determine whether tokens are occur as frequently as these commonly used words either before or after the introduction of a token into the USPTO trademark corpus. Correlations reported are calculated from z-scores for each token each year and the average z-score for the commonly used words each year. Control words used are: other, such, of, in, not,when, and, or, the, a, is, and was.

All tokens selected by the algorithm we propose are associated with some degree of innovation: they are new in the trademark register and their usage is so widespread that sustained and significant adoption of associated products or services in the market place is evident.

In the classification set out here innovations falling into level A align most closely with the notion of innovations as entirely novel and unheard of previously in any domain. Examples included in Table 5 include the spreadsheet, the camcorder, the smartphone, blogging and teleconferencing. This category comprises 18.5% of all tokens in our data.

Level B innovations are based on words for which the 1-gram data reveal some or even constant low-level use in the English language prior to the introduction of the token into the USPTO trademark corpus. On average these terms will have been used up to 200 times per year prior to introduction, which is minimal in comparison to the level of usage recorded for level X innovations reported in Table 10. These tokens are also selected for the significant growth in the usage of the token in the 1-gram data. The usage of these tokens can grow so significantly that they are used over 1000 times more frequently in the period after the token is introduced as a trademark descriptor at USPTO. Due to strict bounds we placed on this second innovation level there are fewer tokens here than in level A.

Level C-E innovations are identified on the basis of the correlation of their usage in the 1-gram data relative to a set of control terms identified by Younes and Reips (2019). These controls capture the frequency with which widely used words appear over time, allowing us to control for whether a token arises at frequencies that diverge from those of these common words before and after the introduction of the token into usage at USPTO. Like level X innovations these tokens capture new forms of products or services that are described using common words. For instance, consider ”statistics” a level D token that becomes less correlated with the control terms after it starts being used at USPTO. Statistics as a field was fully developed by the end of the 1930s. It enters the set of goods and services descriptors used at USPTO in 1990.

The following tables contain information on the average incidence of each token in the trademark token data between 2011 and 2016, the year of introduction, the years after 1940 that the token did not appear in the ngram data as a 1gram, the average number of matches to books before the token is first used to describe trademarks, the growth rate of matches after this period and correlations with the set of control terms before and after the token is first used to describe trademarks.

Table 5: **Top 30 type-A Tokens**

Token	Incidence 2011-16	Year Introduced	**Years no 1-gram**	Matches Before	Matches Growth Rate	Correlation After	Correlation Before
blogging	27279	2000	51	4.4	1278	.85	.17
webcast	6695	1996	49	3.2	132		
webcasting	6695	1996	50	3.3	72	.38	-.8
emailing	6113	1982	22	3.8	346		.36
webinar	6091	2003	61	4	78	.56	.86
microblog	4629	2009	67	10	25	.73	1
microblogging	4629	2009	66	21	44	.8	.99
smartphone	4320	1999	46	5.4	1288	.43	.16
outsourcing	3551	1991	27	52	418	.81	.34
emarketing	2598	1983	37	2.3	25		
terminalling	2598	1983	27	20	.26	-.43	-.12
hepatologic	2194	1998	58				
cartomizer	2161	2011	71				
homepage	1821	1995	33	11	362	.58	.23
dyslipidemia	1528	1997	41	53	21	.69	.74
jeggings	1430	2009	69			-.17	
camcorder	1324	1986	34	125	13	.71	.18
hyperlink	1292	1991	42	5.9	278	.78	.2
hyperlinking	1292	1991	51	3	65	.7	
aromatherapy	1129	1988	24	24	129	.64	.24
downloadable	988	1996	31	31	105	.56	.44
skorts	841	1982	39	2.7	2.8		
webpage	820	1995	33	10	284	.8	.27
teleconferencing	764	1980	25	103	8.2	-.18	-.086
webcam	752	1999	49	4.2	448	.86	.59
spreadsheet	748	1982	25	16	673	.85	-.2
customizable	656	1990	36	16	39	.88	.31
karaoke	653	1987	33	7.6	372	.93	-.29
credentialing	597	1986	27	37	12	.91	.084

Table 6: **Top 30 Type B Tokens**

Token	Incidence 2011-16	Year Introduced	Years no 1-gram	**Matches Before**	**Matches Growth Rate**	Correlation After	Correlation Before
website	68308	1994	4	155	1142	.86	-.0017
dvd	52138	1984	1	14	30	.23	.59
blog	27279	2000	0	58	507	.81	.24
beanie	17452	1985	0	15	38	.78	.33
loadable	15313	1980	1	10	9	.16	.18
wellness	10653	1982	3	23	154	.9	.2
searchable	7866	1994	9	148	24	.71	.23
healthcare	5674	1981	8	50	1064	.92	.16
intranet	4659	1995	17	7	790	.096	-.014
upload	3578	1989	1	25	175	.92	.23
uploading	3578	1989	17	26	75	.91	.14
messaging	3381	1982	16	12	635	.91	.13
laptop	3197	1987	19	16	1450	.89	-.29
desktop	3016	1982	2	93	132	.92	.22
reusable	2885	1981	0	205	16	.89	.33
browser	2808	1991	0	130	133	.85	.29
mousse	2359	1983	0	201	13	.89	.33
mentoring	2143	1989	20	37	563	.82	.052
tasking	1999	1982	0	67	29	.92	.48
voip	1718	1998	9	5.9	61	.13	.12
firewall	1299	1995	0	89	44	.7	.16
troubleshoot	1146	1982	17	29	19	.86	.16
troubleshooting	1146	1982	3	183	13	.9	.49
holdall	1060	1985	0	66	26	.89	.34
biotechnology	978	1981	3	179	127	.19	.22
router	933	1981	0	180	24	.76	.21
chopstick	896	1982	0	28	11	.88	.46
billboards	866	1981	0	133	23	.98	.18
smoothie	840	1993	2	37	71	.65	.61

Table 7: **Top 30 Type C Tokens**

Token	Incidence 2011-16	Year Introduced	Years no 1-gram	Matches Before	Matches Growth Rate	**Correlation** After	Before
consultation	4600	1982	0	42884	1.8	.44	.3
resource	3710	1997	0	34862	7.2	.67	.42
consultancy	3707	1989	0	5017	6.5	-.29	.34
terminal	2598	1983	0	50087	1.2	.5	.68
horticultural	2508	1998	0	8512	-.12	-.16	.56
yoga	2314	1996	0	1170	18	.68	.36
ppc	1961	2008	0	28	1	.57	.42
christian	1888	1994	0	2408	.67	.59	.5
satellite	1561	1980	0	15239	4.3	.48	.32
korean	1416	1984	0	2039	-.94	.67	.043
alzheimer	1240	1986	0	248	-.43	.58	.17
jan	1006	1982	0	825	.43	.68	.59
pelerines	998	1986	8	16	-.22	.2	.38
motorist	980	1993	0	5410	-.35	.56	.33
directories	956	1983	0	2213	3.3	.0075	.35
gussets	948	2002	0	303	-.18	.51	.16
capt	934	1980	0	566	-.077	-.12	.11
hazardous	880	1980	0	8482	3.2	.67	.56
sector	872	1985	0	69273	7.6	.41	.25
crm	835	1993	0	33	2.9	.64	.57
serum	834	1982	0	48359	.47	.35	.64
paas	834	2000	0	24	.57	.18	.42
wap	821	1980	0	81	.4	.29	.31
broadband	737	1994	1	746	19	.58	.31
musculo	708	1980	0	503	-.053	.48	.37
asian	682	1984	0	413	.79	.68	-.031
incubation	622	1995	0	14047	.3	.23	.51
residential	605	1986	0	32699	1.9	.59	.37
korea	573	1982	0	1077	-.83	.63	-.056

Table 8: **Top 30 Type D Tokens**

Token	Incidence 2011-16	Year Introduced	Years no 1-gram	Matches Before	Matches Growth	Correlation After	Correlation Before
flowering	1212	1983	0	29859	.052	.21	.91
cremation	1104	1995	0	3177	2.3	.64	.71
imitations	941	1995	0	5367	.48	.56	.85
fiscal	480	1987	0	23697	3.3	.7	.7
entree	323	1981	0	528	1.2	-.35	.75
curated	285	2011	6	566	9.4	-.28	.79
unworked	273	1986	0	1444	-.22	-.38	.75
collegiate	271	1989	0	3817	.4	.64	.86
employment	242	1999	0	249536	1.1	.46	.71
grape	188	1998	0	6816	2.4	.67	.87
amend	188	1986	0	24991	-.022	.55	.92
keyboarding	185	1981	12	81	3.9	.15	.72
embryo	180	1980	0	16650	.93	.7	.8
statistics	178	1990	0	87893	.94	.55	.7
retouching	168	1983	0	1014	.46	.22	.78
assaying	162	1982	0	1766	-.48	-.45	.79
sclerosis	149	2002	0	6374	3.1	.69	.78
statuettes	138	1997	0	1673	.64	.35	.76
ironmongery	132	1984	0	878	.16	-.18	.79
subscription	122	1980	0	36215	.37	-.68	.88
aided	121	1982	0	24007	.72	.53	.94
western	99	1993	0	102881	1.3	.69	.78
collocation	97	1997	0	1202	2.6	.56	.78
furbishing	95	1989	0	97	-.32	.51	.72
tendering	94	1983	0	3679	1.5	-.28	.88
endocarditis	84	2005	0	3344	2	.62	.72
ballast	82	1991	0	12032	-.042	-.13	.8
abutment	82	1988	0	2234	.019	-.2	.74
frieze	82	1989	0	5746	.52	.26	.86

Table 9: **Top 30 Type E Tokens**

Token	Incidence 2011-16	Year Introduced	Years no 1-gram	Matches Before	Matches Growth Rate	Correlation After	Correlation Before
internet	113407	1993	0	302	375	.8	.032
online	11723	1990	0	2655	85	.88	.22
interactive	6580	1980	0	1074	50	.89	.18
genetic	4218	1981	0	17070	9.5	.94	.28
global	3715	1985	0	11140	50	.92	.21
flop	2948	1982	0	1491	3.8	.95	.27
behavior	2705	1989	0	27939	10	.97	.29
significant	2578	1989	0	160584	5	.92	.34
cognitive	2528	1986	0	8183	28	.96	.24
reaming	2500	1993	0	513	-.14	.77	.13
elderly	2372	1982	0	24001	4.3	.92	.37
radiology	2273	1999	0	1711	1.9	.82	.33
neurological	2124	1984	0	3232	9.4	.94	.27
infrastructure	1749	1988	0	7021	21	.8	.28
inpatient	1729	1996	0	739	13	.81	.37
entrepreneur	1654	1985	0	2653	9.3	.93	.2
membership	1512	1981	0	69379	1.6	.74	.37
operational	1359	1991	0	31842	3.4	.81	.3
threat	1324	1985	0	41854	5.7	.99	.29
content	1258	1985	0	217433	1.6	.97	.39
alerting	1246	1994	1	652	9.4	.95	.31
remediation	1182	1987	9	223	34	.76	.23
personnel	1163	1983	0	47345	2.2	.76	.094
trimmer	1158	1997	0	1037	.082	.86	.34
modeled	1060	1996	0	1275	15	.94	.34
connectivity	1056	1984	1	349	53	.88	.25
protocol	1013	1982	0	4301	13	.91	.28
fantasy	977	1982	0	9159	7.4	.99	.27
sensor	936	1987	0	4364	9.2	.85	.29

Table 10: **Top 30 Type X Tokens**

Token	Incidence 2011-16	Year Introduced	Years no 1-gram	Matches Before	Matches Growth	Correlation After	Correlation Before
family	7958	1995	0	431313	3.3	.99	.74
management	6646	1982	0	171632	5.3	.81	.52
email	6113	1982	0	372	194	.81	.42
analysis	5963	1998	0	353852	3.4	.78	.46
entertainment	5730	1990	0	29535	2.2	.91	.79
marriage	5554	1993	0	167372	1.5	.99	.95
loyalty	4779	1988	0	33071	1.9	.99	.94
addicted	4691	1986	0	3967	2	.98	.93
peer	4617	1996	0	11360	7.5	.97	.73
peering	4617	1996	0	3831	5.9	.9	.86
reality	4450	1987	0	110085	3.7	.97	.74
stream	4121	1991	0	107190	.47	.98	.92
streaming	4121	1991	0	6644	2.7	.95	.9
portal	3565	1982	0	9026	2.1	.94	.81
quiz	3362	2011	0	2238	6.7	.93	.8
urgent	3161	1982	0	45394	.69	.99	.59
spiritual	3047	1987	0	93292	1.4	.99	.85
religious	2939	1990	0	202821	2.3	.93	.93
theft	2753	1997	0	13152	3	.97	.74
down	2737	1991	0	714626	2.2	.95	.97
discovery	2691	1987	0	80926	.91	.98	.95
fitness	2627	1980	0	12099	3.1	.95	.9
define	2566	1981	0	36778	4.3	.96	.51
rental	2452	1981	0	14838	1.6	.87	.92
authentication	2272	1986	0	644	15	.84	.75
outcome	2162	1982	0	37750	5.7	.96	.67
mentor	2143	1989	0	1807	17	.96	.42
venue	2129	1980	0	2873	10	.95	.57
task	1999	1982	0	152090	2	.98	.74

C Further Descriptive Results

This appendix rounds of our discussion of cumulative diffusion curves in Section 3.2. We present four more cases of type-A tokens with comparatively high cumulative trademark counts in 2012. These are tokens for which we found no patents. Futsal is a sport, jeggings are clothes, webinars are related to the webcasts presented in Section 3.2 and bobbleheads are a form of toy. Invention of these things takes place around the time at which the token enters the books in the ngram corpus. Innovations that build on the inventions follow with varying degrees of delay. In the case of jeggings, there is almost no delay at all.

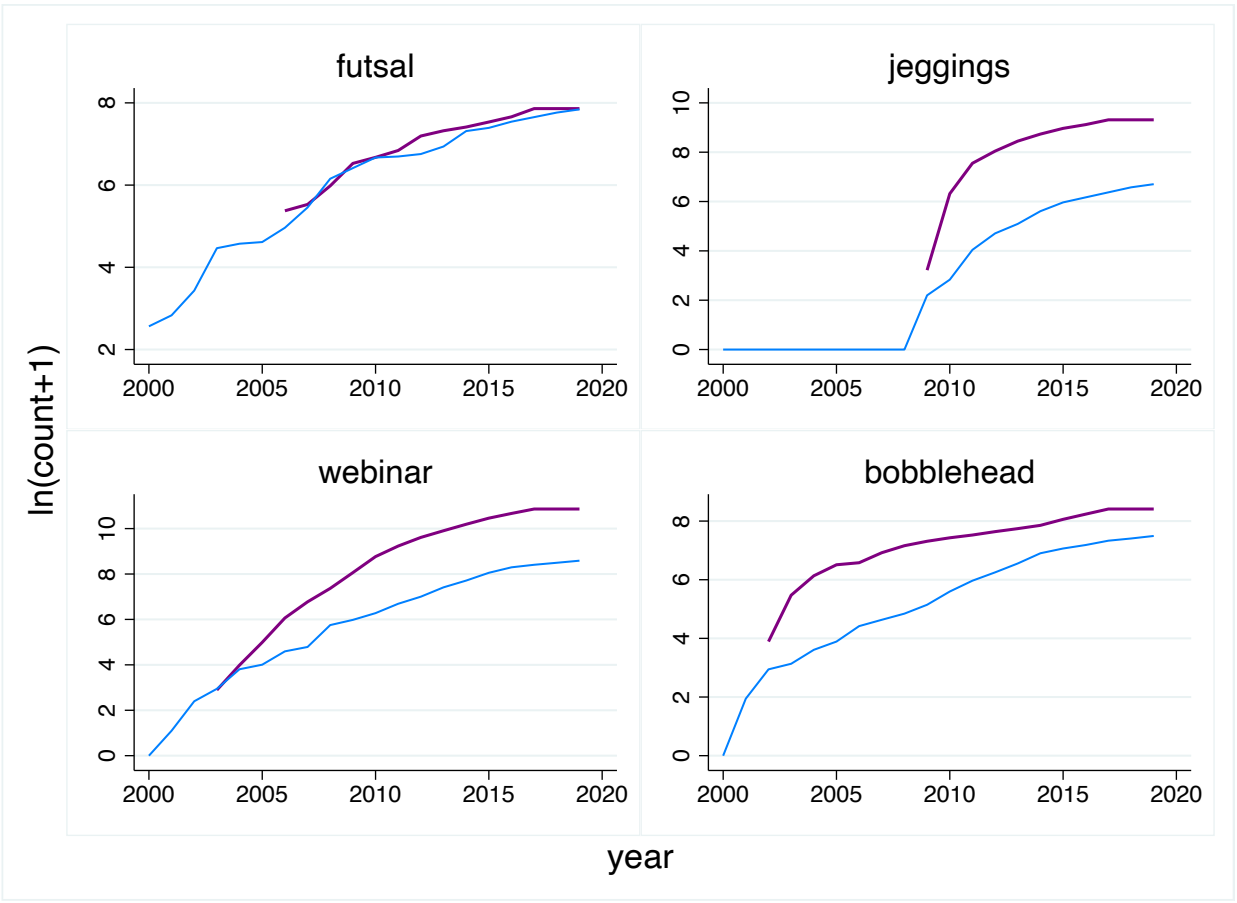

Figure 4: Diffusion of four innovations captured using token counts from ngram (mid-blue line) and trademark token (purple line) data.

Just as with the patentable innovations presented in Section 3.2 we observe a range of diffusion outcomes. While futsal and webinars have grown slowly and continuously, jeggings look more like phenomena that are suddenly very popular, but then cease to spread more widely. The bobblehead is an intermediate case.

These examples demonstrate the wide range of innovations represented in trademark token data, going well beyond the basic dichotomy between goods and services to encompass also games, sports and toys that then give rise to opportunities to sell related products and services.

D Further Analytical Results

This section contains descriptive statistics and analytical results for the whole sample of 9,227 tokens. Theses originate from addresses in 4,295 distinct census tracts and are received by (diffuse to) addresses in 8,659 distinct census tracts. For 14.4% of sending-receiving census tract dyads, the sending and receiving addresses lie within the same census tract.

Balanced panels of census tract and year aggregates of new to the world token introductions and their diffusion are used in our analysis below. Here we identify 14,843 pairs of census tracts having at least one instance of a diffusion from a sending to a receiving census tract in the 32 years between 1981 and 2012. Table 11 sets out descriptive statistics for the resulting panel of tokens of all types.

Table 11: **Descriptive Statistics for All Tokens** (N $= 471,898$)

Variable	**mean**	**sd**	**90th percentile**	**min**	**max**
Diffusion$_{s,r,t}$	0.060	0.692	0	0	231
ln Net Diffusion$_{r,t}$	0.009	0.126	0	0	5.746
Link dummy	0.031	–	0	0	1.000
ln Years linked$_{s,r,t}$	0.878	1.135	2.890	0	3.466
ln Distance$_{s,r,t}$	5.434	2.721	7.829	0	8.541
ln Innovation$_{s,t}$	0.034	0.168	0	0	4.290
ln Innovation$_{s,t-1}$	0.034	0.168	0	0	4.290
ln Innovation$_{r,t-1}$	0.010	0.097	0	0	4.290
ln Net Diffusion$_{s,t-1}$	0.012	0.160	0	0	5.505
Pre-sample Innovation$_s$	0.904	2.866	6	0	26
Pre-sample Innovation$_r$	0.722	2.410	5	0	26

The panel consists of 14,843 dyads of sending/receiving census tracts and covers the years 1981 to 2012.

D.1 First Stage Results

Table 12: First Stage Models: First Diffusion and Net Diffusion

Dependent Variable	First Diffusion		Net Diffusion	
Distance & Periods	in levels	interacted	in levels	interacted
	(1)	(2)	(3)	(4)
$\ln$ Distance$_{s,r,t}$	-0.0003^{***}	-0.0014^{***}	-0.0001	-0.0006^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
$\ln$ Innovation$_{s,t}$	0.2851^{***}	0.2854^{***}	0.0842^{***}	0.0843^{***}
	(0.0109)	(0.0109)	(0.0068)	(0.0068)
$\ln$ Innovation$_{s,t-1}$	0.2273^{***}	0.2276^{***}	0.0527^{***}	0.0528^{***}
	(0.0066)	(0.0066)	(0.0034)	(0.0034)
$\ln$ Net Diffusion$_{r,t-1}$	-0.0167^{***}	-0.0168^{***}		
	(0.0033)	(0.0033)		
$\ln$ Net Diffusion$_{s,t-1}$	-0.0797^{***}	-0.0798^{***}	-0.0220^{***}	-0.0221^{***}
	(0.0030)	(0.0030)	(0.0018)	(0.0018)
$\ln$ Innovation$_{r,t-1}$			0.0164^{**}	0.0167^{**}
			(0.0061)	(0.0061)
$\ln$ Age of link$_t$	-0.0028^{***}	-0.0028^{***}	-0.0005^{***}	-0.0005^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Dummy Service Innovation	-0.0000	-0.0000	0.0020^{***}	0.0020^{***}
	(0.0001)	(0.0001)	(0.0001)	(0.0001)
Pre-sample Innovation$_s$			-0.0105^{***}	-0.0105^{***}
			(0.0004)	(0.0004)
Pre-sample Innovation$_r$	-0.0015^{***}	-0.0015^{***}	-0.0004	-0.0004
	(0.0003)	(0.0003)	(0.0004)	(0.0004)
Period 2 $\times$ $\ln$ Distance$_{s,r,t}$		0.0018^{***}		0.0009^{***}
		(0.0002)		(0.0002)
Period 3 $\times$ $\ln$ Distance$_{s,r,t}$		0.0020^{***}		0.0009^{***}
		(0.0002)		(0.0002)
Constant	0.0202^{***}	0.0202^{***}	0.0135^{***}	0.0135^{***}
	(0.0006)	(0.0006)	(0.0006)	(0.0006)
Observations	471898	471898	471898	471898
R^2	0.1327	0.1329	0.0328	0.0329

[1] Robust standard errors clustered at census tract level in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[2] All models include time fixed effects.

D.2 IV Results

Overall the results set out in Table 13 are remarkably similar to those we discuss in Section 5.2.

Table 13: IV Poisson Models for Diffusion of Innovation - Sample: All Tokens (N=471898)

Model	no IV	Restricted IV		Full IV	
Distance & Periods		in levels	interacted	in levels	interacted
First Diffusion (1/0)	5.2994^{***}	6.2848^{***}	6.1797^{***}	6.0510^{***}	5.9937^{***}
	(0.1238)	(0.6200)	(0.6543)	(0.6202)	(0.6291)
$\ln$ Net Diffusion$_{r,t}$	0.4196^{***}	1.6877^{***}	1.7609^{***}	1.7242^{***}	1.7877^{***}
	(0.0271)	(0.3192)	(0.3466)	(0.3298)	(0.3456)
$\ln$ Distance$_{s,r,t}$	-0.0842^{***}	-0.1311^{***}	-0.0804^{**}	-0.1068^{***}	-0.0633^{*}
	(0.0087)	(0.0173)	(0.0287)	(0.0195)	(0.0283)
$\ln$ Innovation$_{s,t}$	0.6801^{***}			0.2831^{**}	0.2488^{*}
	(0.0785)			(0.0978)	(0.0988)
Pre-sample Innovation $_{s}$	-0.0036	-0.0121	-0.0129	-0.0198	-0.0193
	(0.0045)	(0.0162)	(0.0164)	(0.0163)	(0.0165)
Pre-sample Innovation $_{r}$	0.0017	-0.1550^{***}	-0.1673^{***}	-0.1639^{***}	-0.1742^{***}
	(0.0048)	(0.0417)	(0.0462)	(0.0426)	(0.0454)
Interaction		0.0088^{***}	0.0097^{***}	0.0091^{***}	0.0099^{***}
Pre-sample Innovation		(0.0023)	(0.0027)	(0.0024)	(0.0027)
Constant	-5.0018^{***}	-7.2410^{***}	-7.5809^{***}	-7.4191^{***}	-7.7195^{***}
	(0.1138)	(0.6706)	(0.7984)	(0.6728)	(0.7871)
			Marginal Effects		
$\ln$ Distance$_{\text{pre 1996}}$	-0.0031^{***}		-0.0030^{*}		-0.0023^{*}
	(0.0004)		(0.0012)		(0.0011)
$\ln$ Distance$_{1996-2005}$	-0.0074^{***}		-0.0183^{***}		-0.0158^{***}
	(0.0008)		(0.0043)		(0.0046)
$\ln$ Distance$_{\text{post 2005}}$	-0.0044^{***}		-0.0054^{*}		-0.0040
	(0.0007)		(0.0025)		(0.0025)

[1] Robust standard errors clustered at county dyad level in parentheses: $^{+}$ $p < 0.10$, * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

[2] All models include time fixed effects.

[3] Instruments: $\ln$ Innovation$_{r,t-1}$, $\ln$ Net Diffusion$_{r,t-1}$, $\ln$ Net Diffusion$_{s,t-1}$

Estimating the results set out here was more complex than was the case for the panel restricted to type-A and type-B tokens only. We had to control for pre-sample innovation in sending and receiving locations as well as their interaction.

The pattern of marginal effects we obtain from this analysis is very similar to that presented in Table 3. as there we find persistent negative effects of distance on diffusion of innovations. The strongest effects arise in the period of the Dot com boom. The similarity of effects we see for the restricted sample of type-A and type-B tokens and the full sample suggests that quite similar diffusion mechanisms are at work.