Research Policy 51 (2022) 104418

Contents lists available at ScienceDirect

Research Policy

journal homepage: www.elsevier.com/locate/respol

Short communication

Winner takes all? Tech clusters, population centers, and the spatial transformation of U.S. invention

B. Chattergoon[a,1], W.R. Kerr[a,b,*,1]

[a] Harvard University, Boston, MA, United States
[b] National Bureau of Economic Research, Cambridge, MA, United States

ARTICLE INFO

JEL codes:
O30
O31
O32
O33
O34
R11
R12
L86

Keywords:
Invention
Patents
Innovation
Software
Artificial intelligence
Clusters
Agglomeration

ABSTRACT

U.S. invention has become increasingly concentrated around major tech centers since the 1970s, with implications for how much cities across the country share in concomitant local benefits. Is invention becoming a winner-takes-all race? We explore the rising spatial concentration of patents and identify an underlying stability in their distribution. Software patents have exploded to account for about half of patents today, and these patents are highly concentrated in tech centers. Tech centers also account for a growing share of non-software patents, but the reallocation, by contrast, is entirely from the five largest population centers in 1980. Non-software patenting is stable for most cities, with anchor tenants like universities playing important roles, suggesting the growing concentration of invention may be nearing its end. Immigrant inventors and new businesses aided in the spatial transformation.

One Sentence Summary: The growing concentration of patenting in tech centers masks an important stability in non-software patenting for most U.S. cities.

1. Introduction

The spatial distribution of invention is important for science, business, and policy. Invention builds upon itself, and knowledge spillovers are more localized than other forms of economic interaction.[2] Consequently, tight clusters of innovation form and shape the access of individuals and institutions to important resources necessary for this work.[3] The depths of these local technology pools influence the likelihood of achieving breakthrough inventions that draw frontier industries to a region and the capacity of regions to recombine prior work into novel contributions.[4] The distributional implications of the spatial location of invention are significant and long-lasting, with one study showing children raised in areas lacking invention are less likely to become future inventors (Bell et al., 2019).

U.S. patenting has become much more spatially concentrated around tech clusters like San Francisco and Boston compared to the 1970s, making these places more productive for researchers in terms of their patenting propensity, important for business organization, and central to high-tech startups.[5] Astoundingly, five of the six most valuable public companies *in the world* in 2020 were tech companies headquartered in San Francisco or Seattle. In response, local policy initiatives to boost innovation abound (Chatterji et al., 2014), and 238 U.S. cities bid for Amazon's HQ2. Is invention becoming a winner-takes-all race?

While the growing prominence of tech clusters is important, we show in this note that it is mostly due to two forces: 1) the rise of software patents, which are very concentrated in tech centers, and 2) the

* Corresponding author.
E-mail address: wkerr@hbs.edu (W.R. Kerr).

1 Brad Chattergoon and William Kerr contributed equally to all aspects of the work.
2 Audretsch and Feldman, 1996; Ganguli et al., 2020; Jaffe et al., 1993; Rosenthal and Strange, 2020.
3 Breschi and Lissoni, 2001; Buzard et al., 2017, 2020; Kerr and Kominers, 2015; Stuart and Sorenson, 2003.
4 Duranton, 2007; Duranton and Puga, 2001; Fleming and Sorenson, 2001; Jacobs, 1970; Kerr, 2010; Lin, 2011; Youn et al., 2015.
5 Alcacer and Delgado, 2016; Guzman, 2020; Guzman and Stern, 2020; Moretti, 2019; Verspagen and Schoenmakers, 2004.

https://doi.org/10.1016/j.respol.2021.104418
Received 13 December 2020; Received in revised form 30 August 2021; Accepted 26 October 2021
Available online 30 November 2021

reallocation of non-software patents to tech centers from a few big population centers. These trends mask an important stability in the spatial distribution of non-software patents. We trace part of this stability to dispersed anchor tenants like universities.

Our work is closely linked to Bettencourt et al. (2007) and Balland et al. (2020). Bettencourt et al. (2007) show that patenting activity became increasingly concentrated in U.S. urban areas in the latter decades of the 20th century and that patenting scales at a super-linear rate to city population; Verspagen and Schoenmakers (2004) show similar spatial concentration in multinational patenting in Europe. More recently, Balland et al. (2020) quantify that patenting and related forms of innovation have become increasingly concentrated in larger cities.[6] This increase is linked to the capacity of big cities to conduct more complex processes, and the spatial concentration of invention has been growing since the 1850s.

We contribute to this literature in several ways. Most studies focus on quantifying the macro relationship of patenting to city size, using data spanning small cities like Casper, WY, and Enid, OK to the giants of New York and Los Angeles. Case studies also contemplate competition among tech clusters, such as Saxenian's (1996) account of the migration of semiconductors from Boston to San Francisco. Our contribution is to quantify how much of the rise of tech centers like Boston, Seattle, and San Francisco since the 1970s is due to a shift of patenting from the biggest population centers in 1980 like NYC and LA. The magnitudes are large: the 13.6% reduction from the 1970s to 2015–2019 in the patent share accounted for by the five largest population centers in 1980 is comparable to the combined patenting of the 238 MSAs with the least patenting in 2015–2019.

We also provide new evidence linking this increasing spatial concentration to software/digital inventions, including artificial intelligence. We draw upon algorithms by Bessen and Hunt (2007) and Graham and Vishnubhakat (2013), as well as our own extension of these using machine learning techniques. We measure an enormous increase in the share of patenting that is software related, to account for almost half of patenting today. This type of invention is conducive to spatial concentration and responsible for much of the overall rise in the concentration of inventions. In non-software domains, the distribution of patenting is more stable, although the shift of activity away from large population centers is still evident.[7]

These two trends—the reallocation of patents from a few large population centers to tech clusters and the explosion of software patents—are nuanced and masked in aggregate assessments. The purpose of this research note is to quantify them and raise their profile. We close by exploring their link to factors important for innovation (e.g., universities, immigration). These preliminary explorations are atheoretical and not conclusive, but they hopefully spark interest in follow-on assessments.

2. Patent data

We study micro-records for all utility patents granted by the United States Patent and Trademark Office (USPTO) from January 1976 to December 2020 (Li et al., 2014; Hall et al., 2001). We consider patents with at least one inventor in the United States and locate the work to the modal U.S. city of inventors listed on the patent. We date patents by their application year and consider applications made during 1975 to 2019. Our Online Supplemental Materials describe data preparation in detail.

Defining a tech cluster requires consideration of complementary inputs to patenting like venture capital investment.[8] We follow Kerr and Robert-Nicoud (2020) and Rosenthal and Strange (2020) by using two criteria that reflect the scale and density of local tech activity: 1) the city ranks among the top 15 cities for patents and venture capital investment (the scale of activity) and 2) the city holds shares for patents, venture capital, employment in R&D-intensive sectors, and employment in digital-connected occupations that exceed its population share (the density of activity).

Six metropolitan statistical areas (MSAs[9]) satisfy these scale and density criteria: San Francisco, Boston, Seattle, San Diego, Denver, and Austin. New York and Los Angeles are ambiguous, as the cities hold large scale but fall short on several density requirements. If we relax some requirements, three other candidates are Raleigh-Durham, Minneapolis-St. Paul, and Washington DC, and our Online Supplement Materials discuss robustness of the patterns documented to tech cluster definitions.

In 1980, the ten most populated MSAs were New York City, Los Angeles, Chicago, Philadelphia, Detroit, San Francisco, Washington DC, Dallas-Ft. Worth, Houston, and Boston. San Francisco (#6) and Boston (#10) are two of the identified tech clusters, and the next largest is San Diego at #17 in terms of the 1980 population ranking. Our analysis focuses on the reallocation of patenting from the five largest MSAs in 1980 in terms of population that rank ahead of San Francisco to tech centers.

We identify software patents using algorithms based upon key words in the patent.[10] We build our main results on the algorithm developed by Bessen and Hunt (2007), as it has been commonly used, and we later discuss alternatives.

Bessen and Hunt (2007) state: "Our concept of software patent involves a logic algorithm for processing data that is implemented via stored instructions; that is, the logic is not 'hard-wired.' These instructions could reside on a disk or other storage medium or they could be stored in 'firmware,' that is, a read-only memory, as is typical of embedded software. But we want to exclude inventions that do not use software as part of the invention. For example, some patents reference off-the-shelf software used to determine key parameters of the invention; such uses do not make the patent a software patent."

The Bessen-Hunt algorithm requires that a utility patent description include either the string "software" or the strings "computer" and "program", but it must not contain "antigen" or "antigenic" or "chromatography". The patent title must also not contain "chip" or "semiconductor" or "bus" or "circuit" or "circuitry". The patent titles and grant year of three examples: Apparatus for identifying the type of devices coupled to a data processing system controller (4025906, 1977); Remotely initiated telemetry calling system (5327488, 1992); and Intelligent power cycling of a wireless modem (7308611, 2007).

Software patents have exploded as a share of patenting. During 1975–1979, 2.5% of patents are software related, and this share is 49.9% since 2015. This tremendous growth is due to technological changes to make software widespread and legal changes allowing more intellectual property protection.[11]

3. The spatial transformation of U.S. patenting

Fig. 1 shows annual rates of U.S. patenting for tech clusters and large

[6] In a model of the form y ≈ population^β, the authors estimate β equals 1.54 for published papers, 1.26 for patents, 1.11 for GDP, and 1.04 for employment. Related, they also note a scaling of 1.57 for patents in 'computer hardware and software'.

[7] As we describe further below, the line between software patents and other digitally connected inventions is blurry. The spatial transformation we depict in this note is robust under available definitions, including one for artificial intelligence, but we do not claim that the clustering pattern would be necessarily absent in neighboring domains.

[8] Samila and Sorenson, 2011; Sorenson and Stuart, 2001.

[9] Throughout, we use consolidated MSAs such that San Francisco includes San Jose, Oakland, and so on.

[10] Bessen and Hunt, 2007; Graham and Vishnubhakat, 2013; Hall and MacGarvie, 2010; Layne-Farrar, 2005; Webb, 2019; Webb et al., 2018.

[11] Graham and Vishnubhakat, 2013; Hunt, 2010; Lerner and Seru, 2017.

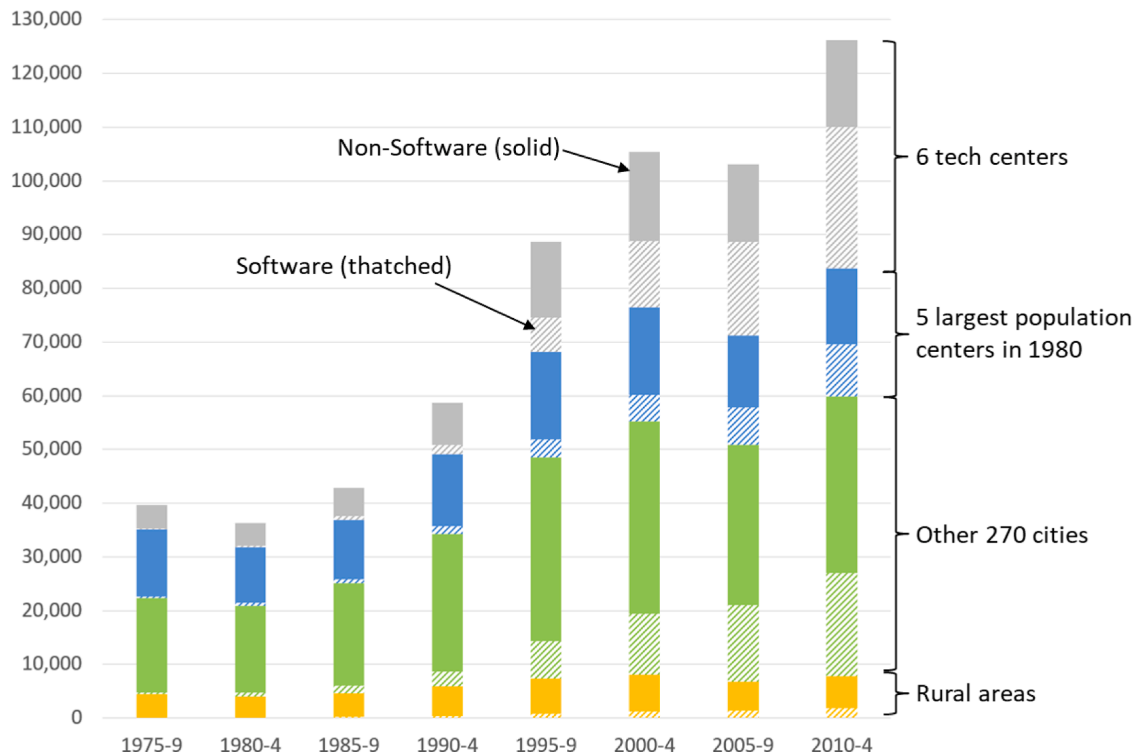


Fig. 1. Figure presents granted U.S. patents by location and software-related. The 6 tech centers are San Francisco, Boston, Seattle, San Diego, Denver, and Austin. The 5 largest cities in 1980 are New York City, Los Angeles, Chicago, Philadelphia, and Detroit. A third group aggregates the remaining 270 metropolitan areas. The thatched portion of each series is software-related, and the solid portion is non-software-related. Patents are dated by their application years, and the final period of 2015-2019 is not shown due to incomplete series with respect to patent counts given future grants will occur.

population centers. Beyond these two groups, we aggregate the remaining 270 MSAs and prepare a fourth group for rural areas. The thatched portion of each series is software-related, and the solid portion is non-software-related. Patents are dated by their application years, and the final period of 2015–2019 is not shown due to incomplete series with respect to patent counts given future grants will occur. The share-based metrics that we focus on for most of this paper are less sensitive to this incomplete process.

The rise of the six tech centers is very stark, and Fig. 2 presents these data in terms of shares. The six tech centers account for 11.3% of patents from 1975 to 1979 but surge to 34.2% for 2015–2019. San Francisco's growth is from 4.6% to 18.4%. While other groups decline in share, the magnitudes and economic importance are different. The five largest population centers show the largest drop, from 32.2% to 18.6%. By contrast, the aggregate decline for the other 270 cities, from 45.5% to 41.0%, is much less. Non-urban areas also decline from 11.0% to 6.1%.

This reallocation is remarkable and has not been documented in prior work. Indeed, because the reallocation is among big cities, this movement of well more than 10% of patents is almost completely orthogonal to the standard elasticity measured across the full city size distribution. In a model of the form patents ≈ population^β, we estimate β=1.313 (0.037) for 1975–1979 and β=1.397 (0.047) for 2015–2019, like prior work. Yet, if we take the patenting that occurs in tech centers and the large population centers for 2015–2019 and re-apportion according to the relative patent shares that were present in 1975–1979, our estimate remains almost identical at β=1.397 (0.042). In other words, the β coefficient is a big vs small city comparison and less sensitive to shifts among bigger cities.[12]

By contrast, an Ellison-Glaeser (EG) metric (Ellison and Glaeser, 1997) calculates the sum of squared deviations between the patenting shares of MSAs compared to their population shares (with a normalization factor). The index is defined as:

$$EG = \frac{\sum_i (s_i - p_i)^2}{1 - \sum_i p_i^2}$$

where s_i is the share of patenting in city i and p_i is the population share. The EG index is well suited for capturing reallocation of activity at the top end of the city size distribution. The EG index has a value of zero if innovation is spread out the same as population; positive values indicate concentration that differs from what one would expect based upon population.

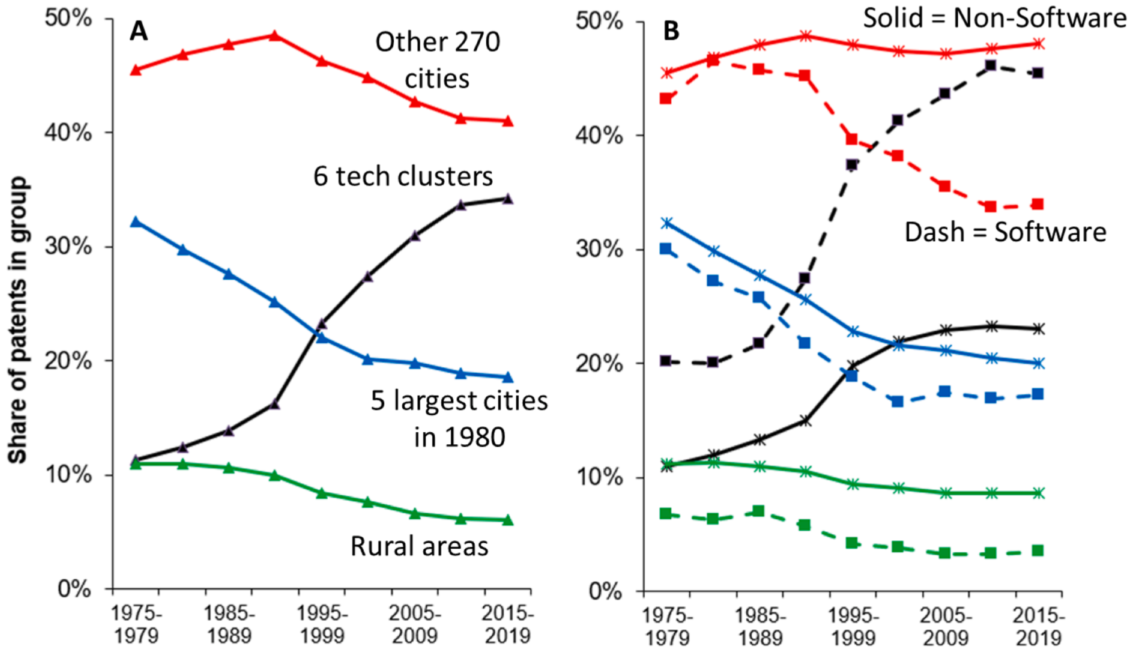
An EG index shows a much stronger response. From a starting value of 0.003 in 1975–1979, the EG increases ten-fold to 0.033 in 2015–2019. This increase is substantial, and if we instead re- re-apportion recent patenting within tech clusters and large population centers according to their relative rates in 1975–1979, our EG index only grows to 0.011. Thus, more than 70% of the rise in the EG index is due to movements among these larger cities.

3.1. Software vs non-software patenting

Figs. 1 and 2 suggest that software patenting is important for our understanding of spatial clustering and tech clusters. Software patents are a significant share of invention in all cities, but they account for well more than half of patents in tech clusters. Panel B in Fig. 2 shows that the tech centers account for 45.4% of software patents after 2015, more than double their starting share of 20.2%. San Francisco again features prominently with 25.8% of software patents filed after 2015. This reallocation pulled from all regions.

Panel B of Fig. 2 shows that tech clusters are also important for non-software patents (solid lines), growing from 11.0% to 23.1% across the period. San Francisco is 11.1%. However, the share for the 270 MSAs

[12] We exclude rural areas from this exercise and the upcoming Ellison-Glaeser calculations.



	All patents			Software patents			Non-Software patents		
	1975-1989	1990-2004	2005-2019	1975-1989	1990-2004	2005-2019	1975-1989	1990-2004	2005-2019
6 tech clusters	12.5	22.3	33.0	20.6	35.4	45.0	12.1	18.9	23.1
5 largest cities in 1980	29.9	22.5	19.1	27.6	19.0	17.2	30.0	23.3	20.6
Other 270 cities	46.7	46.5	41.7	45.1	41.0	34.4	46.8	48.1	47.6
Rural areas	10.9	8.7	6.3	6.7	4.6	3.4	11.2	9.7	8.6

Fig. 2. Figure presents the spatial distribution of granted U.S. patents. The 6 tech centers are San Francisco, Boston, Seattle, San Diego, Denver, and Austin. The 5 largest cities in 1980 are New York City, Los Angeles, Chicago, Philadelphia, and Detroit. A third group aggregates the remaining 270 metropolitan areas. Panel B splits by software vs. non-software patents. Values given in table are averages of the corresponding five-year values in each period.

grows slightly from 45.5% to 48.1%. The shift is instead from the five largest cities in 1980, which fall from 32.3% to 20.0%. These cities have remained mostly prosperous and often hold leading positions in important sectors (e.g., media in Los Angeles, finance in New York). But, while patents continue to increase in a super-linear relationship to city population, invention has become less coupled to the largest cities.[13]

We next separate industrial and university assignees to study agglomeration behavior. Industrial firms have discretion over locations, such as the choice by IBM of how much of its R&D work to conduct in its Yorktown Heights and Albany, NY, labs versus those in Cambridge, MA and San Jose, CA. The creative destruction process also pits new entrants in tech centers against spatially distant incumbents. By contrast, universities are local anchor tenants across the country and mostly constrained from agglomerating.[14] Research universities also rarely go out of business.

Panel A of Fig. 3 displays the EG metric for software and non-software patenting by industrial assignees. Software patenting is more concentrated than non-software patenting, and it has become extremely agglomerated among industrial assignees. As industrial firms account for most patents (85.7% after 2015[15]), their concentration principally shapes the overall concentration of US patenting.

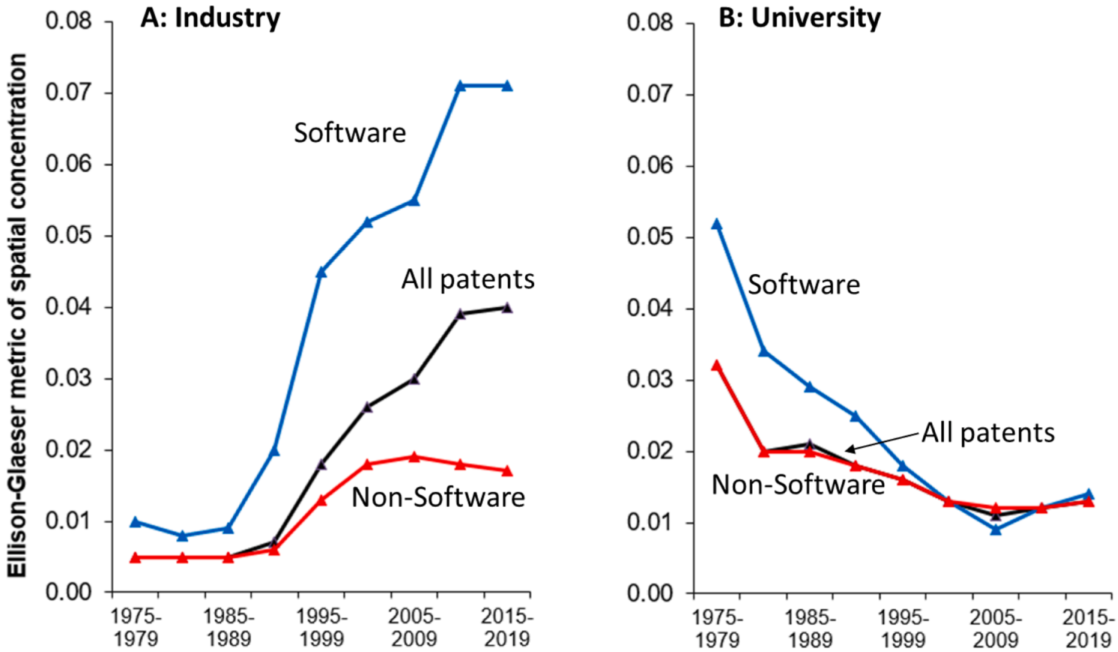
Panel B of Fig. 3 provides a stark contrast with university patenting. While software represents 31.5% of university patents after 2015, their spatial concentration has declined. Concentration levels among non-software patents have also declined.

This recent role of universities in promoting the geographic stability of invention stands in contrast to how universities contributed disproportionately in the 1970s and 1980s for the emergence of software patents in tech centers, especially Boston and San Francisco. After this concentrated start, however, university contributions have been more widespread. The compound annual growth rate of university patenting from 1975 to 2015 is highest in the Other 270 Cities.

[13] Carlino et al., 2007; Fritsch and Wyrwich, 2020; Lerner et al., 2020; Moretti, 2012.

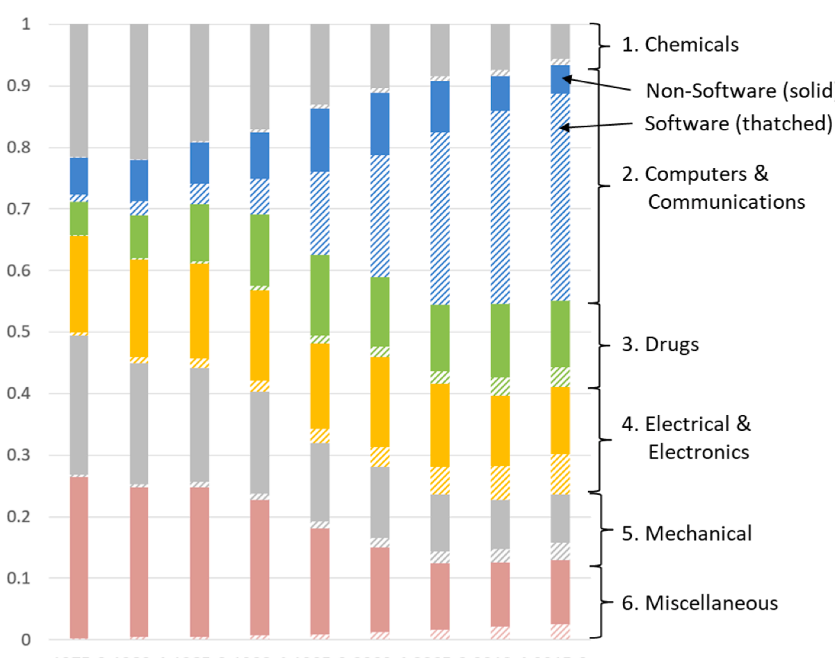
[14] Agrawal and Cockburn, 2003; Agrawal et al., 2014; Berkes and Nencka, 2019; Feldman, 2003; Hausman, 2012; Kantor and Whalley, 2014.

[15] During 1975-1979, approximately 70.2% of patents were made by industrial assignees, 1.1% by universities, 2.8% by government, and 26.0% unassigned. For 2015-2019, these shares were 85.7%, 4.2%, 0.7%, and 9.5%, respectively. Shares can total to more than 100% due to joint assignment of patents across institutions.



	EG: All patents			EG: Industry patents			EG: University patents		
	1975-1989	1990-2004	2005-2019	1975-1989	1990-2004	2005-2019	1975-1989	1990-2004	2005-2019
All	0.003	0.013	0.030	0.005	0.017	0.036	0.024	0.016	0.012
Software	0.008	0.034	0.060	0.009	0.039	0.066	0.038	0.019	0.012
Non-Software	0.003	0.009	0.014	0.005	0.012	0.018	0.024	0.016	0.012

Fig. 3. Figure presents the Ellison-Glaeser index of spatial concentration for granted U.S. patents relative to the population distribution of cities.



2015-2018 shares		
Cat	SW	Non-SW
1	1.09	5.53
2	33.63	4.64
3	3.20	10.87
4	6.56	10.86
5	2.85	7.85
6	2.58	10.33
Total	49.9%	50.1%

Fig. 4. Figure presents the technology composition of granted U.S. patents by the six major NBER categories and whether software-related. The thatched portion of each series is software-related, the solid portion is non-software-related. Shares in table do not precisely sum to 100% due to rounding to two-decimal places.

4. The spread of software patents

In 2011, prominent venture investor Marc Andreessen famously proclaimed "software is eating the world" (Andreessen, 2011). Indeed, Fig. 4 shows that software patenting has expanded beyond its traditional NBER technology categories of computers/communications and electrical/electronics. For example, software patents are 15.8% of patents in chemicals and drugs/medicines. In total for 2012, software patents represent more than a quarter of patents in 24.6% of the 410 United States Patent Classes (USPCs) that are continually present from 1975 to 2012, and more than two-thirds of classes have a greater than 5% software share by 2012.

We can decompose software's growth using the USPC patent class system (which ends in 2012) using the identity:

$$\Delta SW_t = \sum c_{i,t-1}\Delta SW_{i,t} + \sum (SW_{i,t-1} - SW_{t-1})\Delta c_{i,t} + \sum \Delta c_{i,t}\Delta SW_{i,t}$$

where ΔSW_t is the change in the share of software patents between 2012 (t) and 1976 (t-1) and $c_{i,t}$ is USPC class i's share of patents in year t. The first term captures the within-class effect (i.e., software becoming more prevalent as technology classes looked in 1976), the second captures a between-class effect (i.e., classes that were software intensive in 1976 growing more quickly), and the third term represents a cross component (i.e., classes that are becoming more software intensive also growing more quickly).

We calculate that 38.3% of the software patenting growth is from an increased software share holding the 1976 distribution of patent classes constant (the within term), 36.9% from an increase in the class shares holding constant the 1976 software intensity (the between term), and 24.7% from faster growth of classes also correlating with faster software penetration (the cross term). These elements are visible in Fig. 4 as well.

5. Extensions

We discuss here supporting evidence contained in the Online Supplemental Materials.

We defined tech clusters with attention to non-patenting factors (e.g., venture investment). An alternative approach isolates absolute changes in realized patenting growth, which proves informative.[16] The four MSAs that attracted the biggest absolute change in patent counts from 1975 to 2020 are four of our tech clusters (in order): San Francisco, Seattle, Boston, and San Diego. The next three cities in terms of the biggest absolute change in patent counts over the period are Los Angeles, New York City, and Detroit, three of our five large 1980 population centers. Thus, these cities still attracted more patents, as hinted at in Fig. 1, but they lost substantial relative grounds. Indeed, we can replicate our findings with just a focus on the four tech clusters identified with this approach.

Additionally, there is a substantial stagnation and decline in economic might of the Rust Belt during this period. While the major cities in the Rust Belt (such as Buffalo, Cleveland, and Pittsburgh) also lose a substantial share of patenting since the 1970s, this process is distinct. Among our large 1980 population centers, Detroit and, to a lesser extent, Chicago feature among the Rust Belt, but they play a relatively small role in the trends we focus on.

Turning to software definition, a first question focuses on the quality of software patents. Perhaps the explosion in patents in tech centers has been associated with deteriorations in their quality. Using techniques like forward citations,[17] we do not observe any declines in patent quality (software and non-software) for tech centers compared to other locations.

The Bessen and Hunt (BH) technique uses keywords, and a prominent technique by Graham and Vishnubhakat (GV) defines software via patent classes. To evaluate the performance of these approaches, we randomly sampled 1600 patents from NBER Category 2 stratified across eight periods from 1976–79 to 2010–14. Within each period, we sampled 100 BH, 50 GV, and 50 other patents. One patent was sampled twice, and several could not be reliably assigned, resulting in a final sample of 1559 patents. We manually defined 788 (50.5%) of these as software patents.

Both techniques performed reasonably well and had understandable challenges. BH identified 91% of the patents that we classified as software (recall), but only 79% of BH identified patents were ones we identified as software (precision). The parsimonious set of keywords in the BH algorithm performs well in identifying likely software patents, and the algorithm's weakness are the false positives that evade the few negatively selected terms.

The GV algorithm identified 98% of the patents that we classified as software (recall), but only 68% of GV patents identified as software were manually classified as software (precision). The patent class approach achieves a lot by identifying the classes with the most software patents, although 98% will overstate performance if extending sample beyond NBER Category 2. GV's straightforward challenge is that these classes are not exclusive to software patents.

The Online Supplemental Materials show that performance of BH and GV algorithms is best after 1995, increasing on both precision and recall from the 1970s until that point.

Using these 1559 hand-coded patents, we developed a third approach by training a machine learning algorithm. Conceptually, this is an extension of both techniques, giving up the transparency of BH's keywords for a computational approach that particularly bolsters negative selection. The training on many patents in GV classes also benefits from the technology perspectives developed during the examination process. The algorithm is stingier in assignment (88% recall) but has fewer false positives (85% precision).

Fig. 5 shows our key findings with the three techniques. We further incorporate a definition of AI-related patents developed by Giczy et al. (2021). The spatial reallocation of patents that this paper emphasizes are robust across these definitions. There is important scope for further honing patent technology divisions with computation techniques, with this robustness check perhaps being a seed.

5.1. Future research

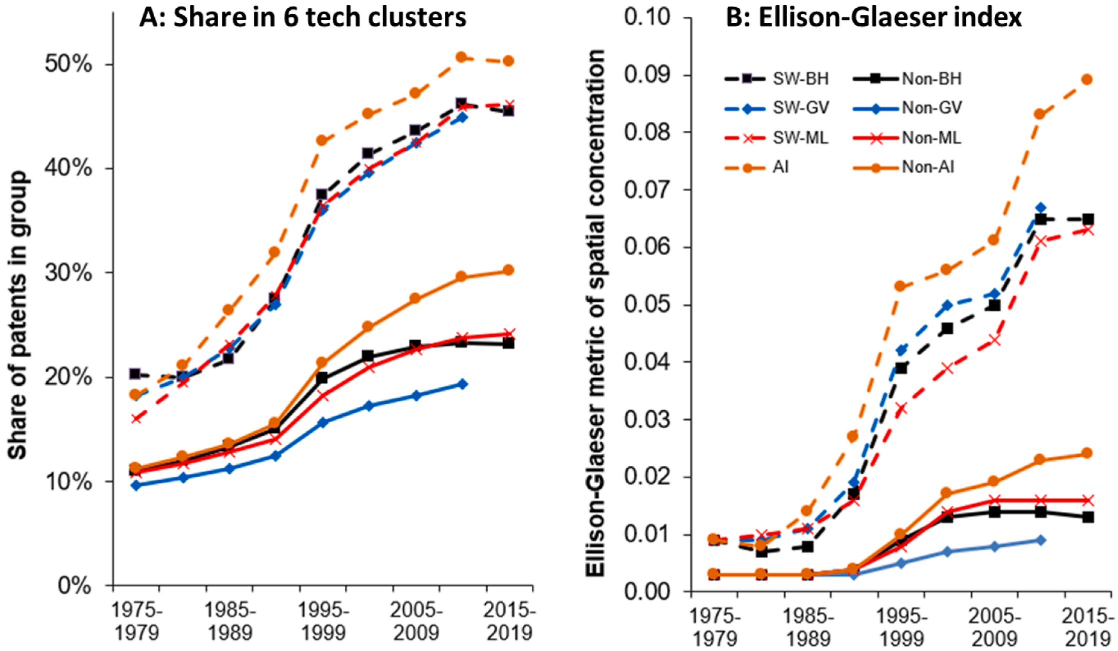
This note has documented a remarkable spatial transformation of patenting due to 1) the rise of software patents, which are very concentrated in tech centers, and 2) the reallocation of non-software patents to tech centers from a few big population centers.

Future research should explore why software patenting rose so much in its spatial concentration. Its higher initial concentration than patenting in traditional technologies (e.g., chemicals, agriculture) is not too surprising, but the subsequent growth in agglomeration deserves attention. Candidate ideas include growing technology complexity (Sorenson et al., 2006; Balland et al., 2020), greater need to participate in tacit knowledge about technology and market trends to be competitive, greater role of venture investment in new software startups, and greater desire for top talent in these fields to be in certain cities.

How this concentration happened is also interesting. In the Online Supplemental Materials, we provide preliminary evidence that software patenting growth in tech centers is facilitated with new businesses coming to the forefront, such as Apple and Microsoft, and less due to shifts in locations of incumbents, such as IBM. We also quantify how the

[16] We thank a referee for this suggestion. In the Online Supplemental Materials, we also discuss the small scope for expanding out the tech cluster definition to include more cities like Raleigh-Durham and Minneapolis-St. Paul.

[17] Harhoff et al., 1999; Hall et al., 2005.



	Share in 6 tech clusters			EG: Software/AI patents		
	1975-1989	1990-2004	2005-2019	1975-1989	1990-2004	2005-2019
SW: Bessen-Hunt	20.6	35.4	45.0	0.008	0.034	0.060
SW: Graham-Vishnubhakat	20.4	34.2	43.7	0.010	0.037	0.060
SW: Machine Learning	19.5	34.7	44.8	0.010	0.029	0.056
AI: Giczy et al.	21.9	39.9	49.3	0.010	0.045	0.078

Fig. 5. Figure presents key outcomes with alternative software and AI definitions. Panel A presents the shares of software and AI related patents in the six tech clusters, as well as the non-software and non-AI shares according to each definition. Panel B presents the Ellison-Glaeser index of spatial concentration for granted US patents relative to the population distribution of cities. BH=Bessen and Hunt (2007) Software; GV: Graham-Vishnubhakat (2013) Software; ML: Machine Learning (authors) Software; and AI: Giczy, Pairolero, and Toole (2021) AI. Values given in table are averages of the corresponding five-year values in each period. GV values in final columns are for 2005-2014.

increased US reliance on immigrant inventors[18] aided the speed of the spatial transition. These two early cuts suggest that the dynamism of the U.S. innovation system, in terms of new firm formation and access to global talent, shaped the spatial transformation.

Hidden behind these trends is an important stability in the spatial distribution of non-software patents. Indeed, to some degree, the spatial transformation of patenting may be ending, excepting for a mechanical effect should software grow as a share of patenting. Fig. 1's trends taper over the last two decades, and the underlying patenting shares of cities are becoming calcified.[19] Despite software's growth across technologies, non-traditional sectors have yet to experience a substantial agglomeration around tech clusters like what transpired in computers/communications and electrical/electronics.

The pandemic raises many ongoing debates about future spatial concentration and tech clusters. Yet, even without the pandemic's emergence, this paper shows the underlying stability of non-software patenting is likely to continue and ensure a broader spatial distribution of innovation. Regional advantages for being a premier location for invention will likely remain the subject of intense local competition,[20] but these spatial dynamics suggest the remarkable recent increases in the concentration of local invention are unlikely to segue into a winner-takes-all race.

Data and materials availability

Data and code are available with online supplement.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

This research was supported by Harvard Business School.

[18] Bernstein et al., 2019; Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Peri et al., 2015; Stephan and Levin, 2001.

[19] To illustrate, a vector of non-software patenting shares for cities in 2015-2019 displays a 0.970 correlation to a similar vector for 1995-1999, whereas the correlation between the vectors for 1995-1999 and 1975-1979 is lower at 0.877.

[20] Chatterji et al., 2014; Gruber and Johnson, 2019; Moretti, 2012; Saxenian, 1996.

Supplementary materials

Supplementary material associated with this article can be found, in the online version, at doi:10.1016/j.respol.2021.104418.

References

Agrawal, A., Cockburn, I., 2003. The anchor tenant hypothesis: exploring the role of large, local, R&D-intensive firms in regional innovation systems. Int. J. Ind. Organiz. 21 (9), 1217–1253.

Agrawal, A., Cockburn, I., Galasso, A., Oettl, A., 2014. Why are some regions more innovative than others? the role of small firms in the presence of large labs. J. Urban Econ. 81 (1), 149–165.

Alcacer, J., Delgado, M., 2016. Spatial organization of firms and location choices through the value chain. Manage. Sci. 62 (11), 3213–3234.

Andreessen, M., 2011. Why software is eating the world. Wall Street J. https://www.wsj.com/articles/SB10001424053111903480904576512250915629460.

Audretsch, D.B., Feldman, M.P., 1996. R&D spillovers and the geography of innovation and production. Am. Econ. Rev. 86 (3), 630–640.

Balland, P.A., Jara-Figueroa, C., Petralia, S.G., Steijn, M.P.A., Rigby, D.L., Hidalgo, C.A., 2020. Complex economic activities concentrate in large cities. Nat. Hum. Behav. 4, 248–254. https://doi.org/10.1038/s41562-019-0803-3.

Bell, A., Chetty, R., Jaravel, X., Petkova, N., Van Reenen, J., 2019. Who becomes an inventor in America? The importance of exposure to innovation. Q. J. Econ. 134 (2), 647–713.

Berkes, E., P. Nencka. (2019). 'Novel' ideas: the effects of Carnegie libraries on innovative activities. Working Paper.

Bernstein, S., Diamond, R., McQuade, T., Pousada, B. (2019). The contribution of high-skilled immigrants to innovation in the United States. Working Paper.

Bessen, J., Hunt, R.M., 2007. An empirical look at software patents. J. Econ. Manage. Strat. 16 (1), 157–189.

Bettencourt, L., Lobo, J., Strumsky, D., 2007. Invention in the city: increasing returns to patenting as a scaling function of metropolitan size. Res. Policy 36, 107–120. https://doi.org/10.1016/j.respol.2006.09.026.

Breschi, S., Lissoni, F., 2001. Knowledge spillovers and local innovation systems: a critical survey. Ind. Corporate Change 10 (4), 975–1005.

Buzard, K., Carlino, G., Hunt, R.M., Carr, J., Smith, T., 2017. The agglomeration of American R&D labs. J. Urban Econ. 101, 14–26.

Buzard, K., Carlino, G., Hunt, R.M., Carr, J., Smith, T., 2020. Localized knowledge spillovers: evidence from the spatial clustering of R&D labs and patent citations. Reg. Sci. Urban. Econ. 81, 103490.

Carlino, G.A., Chatterjee, S., Hunt, R.M., 2007. Urban density and the rate of invention. J. Urban Econ. 61 (3), 389–419.

Chatterji, A., Glaeser, E., Kerr, W., 2014. Clusters of entrepreneurship and innovation. Innov. Policy Econ. 14, 129–166.

Duranton, G., 2007. Urban evolutions: the fast, the slow, and the still. Am. Econ. Rev. 97 (1), 197–221.

Duranton, G., Puga, D., 2001. Nursery cities: urban diversity, process innovation, and the life cycle of products. Am. Econ. Rev. 91 (5), 1454–1477.

Ellison, G., Glaeser, E., 1997. Geographic concentration in U.S. manufacturing industries: a dartboard approach. J. Polit. Econ. 105 (5), 889–927.

Feldman, M., 2003. The locational dynamics of the US biotech industry: knowledge externalities and the anchor hypothesis. Ind. Innov. 10 (3), 311–329.

Fleming, L., Sorenson, O., 2001. Technology as a complex adaptive system: evidence from patent data. Res. Policy 30 (7), 1019–1039.

Fritsch, M., Wyrwich, M. (2020). Is innovation (increasingly) concentrated in large cities? an international comparison. Working paper.

Ganguli, I., Lin, J., Reynolds, N., 2020. The paper trail of knowledge spillovers: evidence from patent interferences. Am. Econ. J. 12 (2), 278–302.

Giczy, A., Pairolero, N., Toole, A. (2021). Identifying artificial intelligence (AI) invention: a novel AI patent dataset. USPTO Economic Working Paper No. 2021-2.

Graham, S., Vishnubhakat, S., 2013. Of smart phone wars and software patents. J. Econ. Perspect. 27 (1), 67–86.

Gruber, J., Johnson, S., 2019. Jump-Starting America: How Breakthrough Science Can Revive Economic Growth and the American Dream. Public Affairs, New York.

Guzman, J. (2020). Go west young firm: the value of entrepreneurial migration for startups and their founders. Working Paper.

Guzman, J., Stern, S., 2020. The state of American entrepreneurship: new estimates of the quality and quantity of entrepreneurship for 32 US states, 1988–2014. Am. Econ. J. forthcoming.

Hall, B., Jaffe, A., Trajtenberg, M. (2001). The NBER patent citation data file: lessons, insights and methodological tools. NBER Working Paper 8498.

Hall, B.H., Jaffe, A., Trajtenberg, M., 2005. Market value and patent citations. RAND J. Econ. 36 (1), 16–38.

Hall, B., MacGarvie, M., 2010. The private value of software patents. Res. Policy 39 (7), 994–1009.

Harhoff, D., Narin, F., Scherer, F.M., Vopel, K., 1999. Citation frequency and the value of patented inventions. Rev. Econ. Stat. 81 (3), 511–515.

Hausman, N. (2012). University innovation, local economic growth, and entrepreneurship. Center for Economic Studies Paper 12–10.

Hunt, J., Gauthier-Loiselle, M., 2010. How much does immigration boost innovation? Am. Econ. J.: Macroeconomics 2 (2), 31–56.

Hunt, R., 2010. Business method patents and U.S. financial services. Contemp. Econ. Policy 28, 322–352.

Jacobs, J., 1970. The Economy of Cities. Vintage Books, New York.

Jaffe, A.B., Trajtenberg, M., Henderson, R., 1993. Geographic localization of knowledge spillovers as evidenced by patent citations. Q. J. Econ. 108 (3), 577–598.

Kantor, S., Whalley, A., 2014. Knowledge spillovers from research universities: evidence from endowment value shocks. Rev. Econ. Stat. 96 (1), 171–188.

Kerr, W.R., 2010. Breakthrough inventions and migrating clusters of innovation. J. Urban Econ. 67 (1), 46–60.

Kerr, W.R., Kominers, S.D., 2015. Agglomerative forces and cluster shapes. Rev. Econ. Stat. 97 (4), 877–899.

Kerr, W.R., Lincoln, W., 2010. The supply side of innovation: H-1B visa reforms and U.S. ethnic invention. J. Labor Econ. 28 (3), 473–508.

Kerr, W.R., Robert-Nicoud, F., 2020. Tech clusters. J. Econ. Perspect. 34 (3), 50–76.

Layne-Farrar, A., 2005. Defining software patents: a research field guide. SSRN Electron. J. https://doi.org/10.2139/ssrn.1818025.

Lerner, J., Seru, A. (2017). The use and misuse of patent data: issues for corporate finance and beyond. NBER Working Paper 24053.

Lerner, J., Seru, A., Short, N., Sun, Y. (2020). Financial innovation in the 21st century: evidence from U.S. patenting. Working Paper.

Li, G.C., Lai, R., D'Amour, A., Doolin, D.M., Sun, Y., Torvik, V.I., Amy, Z.Y., Fleming, L., 2014. Disambiguation and co-authorship networks of the US patent inventor database (1975–2010). Res. Policy 43 (6), 941–955.

Lin, J., 2011. Technological adaptation, cities, and new work. Rev. Econ. Stat. 93 (2), 554–574.

Moretti, E., 2012. The New Geography of Jobs. Houghton Mifflin Harcourt, New York.

Moretti, E. (2019). The effect of high-tech clusters on the productivity of top inventors. NBER Working Paper 26270.

Peri, G., Shih, K., Sparber, C., 2015. STEM workers, H-1B visas and productivity in U.S. cities. J. Labor Econ. 33 (S1), S225–S255.

Rosenthal, S., Strange, W., 2020. How close is close? the spatial reach of agglomeration economies. J. Econ. Perspect. 34 (3), 27–49.

Samila, S., Sorenson, O., 2011. Venture capital, entrepreneurship, and economic growth. Rev. Econ. Stat. 93 (1), 338–349.

Saxenian, A.L., 1996. Regional Advantage: Culture and Competition in Silicon Valley and Route 128. Harvard University Press, Cambridge, MA.

Sorenson, O., Rivkin, J.W., Fleming, L., 2006. Complexity, networks and knowledge flow. Res. Policy 35 (7), 994–1017.

Sorenson, O., Stuart, T.E., 2001. Syndication networks and the spatial distribution of venture capital investments. Am. J. Sociol. 106 (6), 1546–1588.

Stephan, P., Levin, S., 2001. Exceptional contributions to US science by the foreign-born and foreign-educated. Popul. Res. Policy Rev. 20 (1), 59–79.

Stuart, T., Sorenson, O., 2003. The geography of opportunity: spatial heterogeneity in founding rates and the performance of biotechnology firms. Res. Policy 32 (2), 229–253.

Webb, M. (2019). The impact of artificial intelligence on the labor market. Working Paper.

Webb, M., Short, N., Bloom, N., Lerner, J. (2018). Some facts of high-tech patenting. NBER Working Paper 24793.

Verspagen, B., Schoenmakers, W., 2004. The spatial dimension of patenting by multinational firms in Europe. J. Econ. Geogr. 4, 23–42.

Youn, H., Strumsky, D., Bettencourt, L.M.A., Lobo, J., 2015. Invention as a combinatorial process: evidence from US patents. J. R. Soc., Interface 12 (106), 20150272.