Journal of Economic Geography **16** (2016) pp. 1259–1278
Advance Access Published on 23 October 2016
doi:10.1093/jeg/lbw035

River deep, mountain high: of long run knowledge trajectories within and between innovation clusters[1]

Önder Nomaler and Bart Verspagen**,†*

*Eindhoven University of Technology, ECIS The Netherlands, and UNU-MERIT, Maastricht, The Netherlands
**UNU-MERIT and Department of Economics, Maastricht University, PO Box 616, Maastricht, The Netherlands
†Corresponding author: Eindhoven University of Technology, ECIS The Netherlands, and UNU-MERIT, Maastricht, The Netherlands. *email* <b.verspagen@maastrichtuniversity.nl>

Abstract

We bring together the topics of geographical clusters and technological trajectories, and shift the focus of the analysis of regional innovation to main technological trends rather than firms. We define a number of inventive clusters in the US space and show that long chains of citations mostly take place between these clusters. This is reminiscent of the idea of global pipelines of knowledge transfer that is found in the geographical literature. The deep citations are used to identify technological trajectories, which are the main directions along which incremental technological progress accumulates into larger changes. While the origin and destination of these trajectories are concentrated in space, the intermediate nodes travel long distances and cover many locations across the globe. We conclude by calling for more theoretical and empirical attention to the 'deep rivers' that connect the 'high mountains' of local knowledge production.

Keywords: Patent citations; regional concentration of inventive activities; technological trajectories
JEL classifications: O33, O31, R11
Date submitted: 30 December 2015 **Date accepted:** 19 September 2016

1. Introduction

The concentration of specific kinds of economic activity in small and confined geographical spaces is a long-lasting theme in the economics, business and geographical literature (cf. Marshall, 1890; Porter, 2000; Dicken and Lloyd, 1990). The development of new knowledge motivated by economic incentives is one of these activities that seem to be geographically clustered in a strong way (e.g. Storper, 1993). The literature on clusters of innovative activity suggests that a knowledge-based theory is needed to explain the more general trend of industry agglomeration (Malmberg and Maskell,

1 We thank participants at the 9th Summer Academy on 'Innovation and Uncertainty' in Jena, Germany, 26 July–8 August 2015, participants at the workshop on 'Evolutionary approaches informing research on entrepreneurship and regional development', Gothenburg, 9–11 December 2015, for comments to a presentation of the results presented in this paper, and two anonymous referees and the editor for very useful comments on a first draft.

2002), or that the knowledge creating capacity of local clusters is decisive for growth and the long-run survival of the cluster (Bathelt, 2005).

Traditionally, the explanation for the tendency clustering has been sought in the broad concept of agglomeration economies, i.e. the idea that by locating close together, firms may reduce production costs. Agglomeration economies comprise a wide variety of factors (Dicken and Lloyd, 1990), including the availability of raw materials, the availability of a pool of specialized labour, the possibilities for an advanced division of labour, and, specifically for innovation clusters, knowledge spillovers that are bounded by distance (Jaffe et al., 1993). The role of social networks and entrepreneurship has also been stressed as a factor explaining clustering of innovation activities (Sorenson, 2003).

An important topic in the literature on innovation clusters is which specific aspects of the nature of knowledge give rise to clustering. The tacit versus codified nature of knowledge may explain an important part (Gertler, 2003). The part of knowledge that cannot be codified, i.e. which is tacit, can most easily be transferred over small distances (by face-to-face contact). This gives rise to agglomeration economies in knowledge transfer and knowledge spillovers. The degree of complexity of knowledge has also been suggested as a factor in innovation clustering (Sorenson et al., 2006).

The concentration of innovation activities in confined spaces may also lead to a particular interactive structure between firms and other actors (such as research institutes, universities, regional policymakers including investment agencies) that acts as a system. This introduces the notion of a regional innovation system (e.g. Morgan, 2004; Lau and Lo, 2015). The crucial aspect of these systems is that institutions (in the widest possible definition, i.e. including formal rules of the game as well as informal customs and habits) partly determine the performance of the system. This implies that there is a meso-structure to regional innovation systems, which surpasses the level of decision-making in individual firms.

At an abstract level, what these approaches have in common is that they stress the importance of knowledge flows for the generation of new knowledge. Scientists and engineers that are engaged in the inventive process use existing knowledge to generate new ideas. In other words, knowledge is cumulative. The combination of the cumulative nature of knowledge and the localized nature of knowledge flows gives rise to the existence of geographical clusters of knowledge production.

But knowledge flows are not exclusively local. This recognition gave rise to the idea that firms use a combination of 'local buzz and global pipelines' (Bathelt et al., 2004) for sourcing new knowledge as an input into their own innovation activities. The local buzz that firms find inside the cluster consists of easily accessible knowledge resources that are 'just there', like Marshall's (1890) idea that the mysteries of trade are no longer mysteries but are 'in the air'. Local buzz is the typical factor behind the success of Silicon Valley, where engineers, entrepreneurs and venture capitalists meet in bars and restaurants (Saxenian, 1994). Global pipelines, on the other hand, are consciously constructed and managed gateways to knowledge in different (long distance) localities. Strategic alliances (Owen-Smith and Powel, 2004) have been analyzed as a tool for building global pipelines. Bathelt et al. (2004) argue that the co-existence of a high level of local buzz and access to many global pipelines provides the best environment for economic success, both for individual firms, and for cluster regions. Bathelt and Glückler (2011) discuss the broader consequences of local buzz and global pipelines for the geography of innovation.

The literature on innovation clusters, regional innovation systems or local buzz versus global pipelines, takes the firm as the key unit for analyzing knowledge dynamics. The knowledge that this literature is interested in is economically motivated, and this makes firm decisions a logical starting point for the analysis. The novel contribution of the current article is that we want to shift this focus from the firm to technology itself. We are not so much interested in the evolution of the firm population that invests in knowledge and that operates in local clusters, but rather in the way that technology itself evolves, and how this evolution interacts with space. The way in which we undertake this endeavor is to draw on an existing literature that conceptualizes technological evolution as trajectories, and which quantifies these trajectories by using patent citations data. We will briefly summarize the main ideas in this literature below.

Technological trajectories are accumulated chains of incremental innovations (we use patent data to measure these innovations) that display the dominant long-run developments in technology. Examples of technological trajectories include Moore's law, which defines technological progress in personal computers, or the specific types of internal combustion engines used in motor cars. The idea of technological trajectories is important for analyzing the main avenues along which technological change has an impact on the economy and society at large, including the firms that are the subject of the geographical literatures that were briefly referred to above. But when adopting a firm perspective, as much of the literature does, the main directions of technological change remain largely obscure.

The results of our analysis point to the importance of knowledge flows between local innovative clusters, rather than within clusters, for the development of technological trajectories. Although we confirm the strongly concentrated nature of knowledge generation activities, and of patent citations (which are the main unit of observation in our analysis), our results show that the accumulated knowledge flows between rather than within clusters are responsible for the main directions of technological change (the trajectories). This emphasis on between-clusters flows is reminiscent of the idea of global pipelines, but it differs in one crucial aspect. Whereas the global pipelines that firms use to access knowledge from far-away locations are consciously managed and constructed, the trajectories that we find emerge from collective rather than individual action. No single firm exclusively shapes a technological trajectory (see, e.g. the empirical evidence in Verspagen, 2007).

The research here is a novel combination of existing research traditions that have so far not been combined at all. One overlap that can be observed between these literatures is the use of data on patents and patent citations, which is a (small) part of the geographical literature (e.g. Sorenson et al., 2006; Jaffe et al., 1993), and dominates the branch of the trajectories literature that attempt to quantify the main concepts in this tradition (e.g. Verspagen, 2007, Mina et al., 2007; Martinelli and Nomaler, 2014). Another overlap lies in the use of formal network methods, e.g. Owen-Smith and Powell (2004) in the analysis of local buzz and global pipelines, while a specific form of network theory is the bread and butter of quantifying technological trajectories.

Otherwise, the underlying units of analysis (relationships between firms and other organization on the one hand vs. relationships between inventions on the other hand), and the disciplinary backgrounds (geography, business studies and economics on the one hand versus economic history and technology analysis on the other hand) are very different between the fields of literature that we try to bridge. Fitting with the early stage of such an ambitious combination of literatures, our analysis will be mainly explorative in nature. We will try to operationalize various concepts and ideas from both literatures and bring them together in an empirical overview of the trends that are observed in our database. We leave a fuller development of truly integrative conceptualizations to a later stage, which we hope will be forthcoming on the basis of the interest in the empirical facts that we provide.

Ultimately, what these empirical facts suggest is that the geographical concentration of patent citations that has been an important topic of the literature so far, is typical of individual small technological steps, while the main directions of technological change (trajectories) that are comprised of many cumulative incremental steps have a much wider spatial reach than is suggested by the analysis of patent citations by economic geographers. The underlying reason for this broader geographical reach lies in between-cluster knowledge flows. The collective nature of these between-cluster flows adds a new dimension to the understanding of global pipelines that are analyzed in a part of the geographical literature.

The rest of the article is organized as follows. In Section 2, we provide a short overview of the relevant literature on technological trajectories, including a discussion of the indicators used (patents and patent citations). Section 3 introduces the database, and Section 4 presents the methods used. Section 5 is the first one where we present novel empirical results. Here, 35 inventive clusters in the USA are presented, which will form the main unit of analysis in the remainder of the article. Our analysis focuses exclusively on the USA, because for this country we have data available that are broken down to the relevant geographical unit (counties). Section 6 looks at the geographical distribution of citations, both direct (which is the usual indicator of knowledge flow between localities), and so-called deep citations, which are our way of identifying technological trajectories. Section 7 looks in-depth at some of these technological trajectories, in this case between the largest of the 35 US inventive clusters. The concluding Section 8 discusses how the evidence provided in our analysis has implications for the spatial nature of knowledge flows and knowledge production, and suggests directions for future research.

2. Technological trajectories and patent citations

The novelty of our research lays in the application of the idea of technological trajectories to the idea of innovation clusters, and in particular the quantification of a geographical dimension of technological trajectories. The notion of technological trajectories stems from a number of authors (in particular Dosi, 1982; Sahal, 1981) who analyze the history of technology from a strongly economic perspective. The central idea is that the economic impact of innovation takes place through a combined process of radical breakthroughs, incremental innovations and diffusion. This generates ‘technological paradigms’ and ‘technological trajectories’ (Dosi, 1982). By a technological paradigm, Dosi refers to a ‘model and pattern of solution of selected technological problems, based on selected principles from the natural science and on selected material technologies’. A paradigm is the set of technological opportunities that emerges from a radical breakthrough, such as the application of steam power to industrial processes, or the notion of mechanized calculations based on binary logic.

The paradigm develops along a number of specific trajectories, which are accumulations of incremental improvements of a basic design. These incremental innovations are endogenous reactions to the specific circumstances in which the technology develops. For example, when labor costs are high and an important part of total production costs, the trajectory will likely take a labor-saving nature. This endogeneity of the main technological trends also implies that technological trajectories may lock-in to a particular direction, and ignore technological possibilities that lie further away in technological space (e.g. Arthur, 2014). As an evolutionary process, technological change does not optimize globally, but adapts to local circumstances.

An example of a technological trajectory is the famous 'Moore's Law' that describes the technological development of microprocessors (the law states that the number of transistors in a single integrated circuit doubles every two years). How different technological trajectories co-exist within the same paradigm is illustrated well by the example of steam engines. In one particular environment, trains, a trajectory of lean but powerful high-pressure engines emerged, while in the case of Cornish metal ore mines a completely different trajectory of very large engines with relatively low pressure developed (Nuvolari and Verspagen, 2009). In summary, a technological paradigm is a set of radical breakthroughs that defines developments in the techno-economic domain for the long run, while the technological trajectory adapts the paradigm to local circumstances through a series of cumulative and incremental innovations. Although made up of incremental steps, technological trajectories represent big changes over long periods of time. It is this kind of change that we are interested in here.

Verspagen (2007) and Mina et al. (2007), based on Hummon and Doreian (1989) pioneered a method to map technological trajectories using patent citations. Their approach focuses on a small pre-defined field (fuel cells in the case of Verspagen, 2007; and medical technology in the case of Mina et al., 2007), for which it identifies a number of citation paths that capture the largest amount of knowledge flows in the field. The current article applies the same method, with further developments, to the phenomenon of spatial knowledge flows. Thus, our analysis makes the novel combination of mapping trajectories of knowledge in technological space, with trajectories of knowledge in geographical space. Moreover, instead of analyzing the trajectories in a single technology field (which is the norm in the literature), we look at a much larger patent dataset that covers all trajectories in all technology fields in the period under consideration.

The quantitative analysis of technological trajectories is mostly done with patent data, in particular with patent citations. These data are also used in the geographical literature on the concentration of patenting activities. Jaffe et al. (1993) found that in the USA, distance is inversely related to the probability that two patents are linked by a citation. They control for a range of factors such as technology class and time, by matching actual citation pairs by pairs of patents that are similar in terms of these other variables, but do not cite each other, and then find that distance across the two patents in the pair is smaller in the group of patent citations. This result was confirmed in many follow-up studies, including other geographical areas, e.g. Maurseth and Verspagen (2002) and Bottazzi and Peri (2003) for Europe.

The use of patent statistics in geographical analysis requires careful interpretation. Griliches (1990) discusses many aspects of patents as indicators. Perhaps the most crucial aspect of his discussion is the fact that patents are indicators of invention rather than innovation. They show the technical possibilities, but do not guarantee commercial relevance. In fact, many patents that are granted are not used commercially (Giura et al., 2007). This is more of a limitation for studies that aim at analyzing the economic success of innovative firms, than it is for our study. As already explained, our emphasis is on technological trajectories, which represent the main directions of technology. Inventions are the basic unit of these trajectories, and whether or not these inventions are actually commercialized is of secondary importance as compared to a research design where the emphasis is on firms. That individual patents are the incremental steps that accumulate into a trajectory is consistent with the finding by Giura et al. (2007) and Gambardella et al. (2008), who show that most patents have small economic value, and provide small technological steps, with only a few very infrequent outlier patents representing radical change and large economic value.

Griliches (1990) also stresses that the extent to which firms patent their inventions differs greatly between industries. For example, in pharmaceuticals, patents are of crucial importance, because a product that is not patented can easily be imitated, at lower costs, by competitors. In other industries, such as machinery, patents are less important, because competitiveness depends more on factors that are not described in the patent itself. Also, some inventions cannot (or could not) be patented, such as software in some jurisdictions. As a result of this, patents are very common in some industries, and not in others. This does affect our analysis, as the geographical patterns of patented inventions may differ from non-patented innovations.

This literature has also addressed the issue of whether patent citations can actually be seen as a measure of spillovers or knowledge flows. This is, again, mostly relevant for studies that take the firm (or innovative efforts by the firm) as the main object of study. In many of those studies, for example, when analyzing the nature and causes of clustering of innovative efforts, the factor of interest are the flows that firms receive, and use to generate new knowledge. The implicit assumption is that patent citations indicate such flows, from the patent that is cited to the patent that is citing. The fact that many citations are added by patent examiners (instead of by the inventor herself), may be a reason in itself why the citation does not indicate an actual technology flow, as it seems to suggest that the inventor of the citing patent did not know the cited patent. This has been investigated, for example, by Thompson (2006) and Criscuolo and Verspagen (2008). The conclusion seems to be that although citations are noisy indicators, the conclusion of a geographical bias in knowledge flows stands even when only inventor-citations are used.

For our purposes, when analyzing technological trajectories, patent citations will be used as indicators of technological relatedness. The citing patent is related to the cited patent, even if the inventor did not make the citation, because it is the job of the patent examiner to judge the novelty of the patent, and citations are used for that purpose. It is this kind of technological relatedness that the trajectory approach uses to map the main directions of technological developments. We will use the term technology flows for this, even though we do not imply that a flow between two firms (or inventors) has necessarily taken place.

3. Data

We use the OECD REGPAT database, which contains patent-level information, with geographical information about the inventors and applicants of the patent. The patents in the REGPAT database are patents issued by the European Patent Office (EPO), or filed under the so-called Patent Cooperation Treaty (PCT), which allows just one application at one of the participating offices, and get patents in multiple jurisdictions. We focus on the USA, in particular the part of the USA located on the main North American continent, with the exception of Alaska. The geographical entities used in the REGPAT database are always administrative regions. In the USA, the regions are counties. In other nations, the geographical entity tends to be much larger, which is why we focus on the USA.

We generally count patents in a fractional way, i.e. when there is more than one inventor, the patent is assigned to all geographical entities (regions) that the inventors come from, using weights that are proportional to the number of times a region appears on the inventor list. The same fractional procedure is applied to citations, which have a citing and a cited inventor list.

We divide the total period for which we have data, which is 1973–2012, into sub-periods of 3 years, and focus on the time span that starts with 1986–1988 and ends with 2004–2006. The EPO only started in 1979, and had relatively few patent applications in the early years (before 1986). From 2007 to 2009, the number of patents declines due to a backlog in processing (our dates are priority dates, as close as possible to the date of invention). These trends and other basic information about the data will be illustrated and discussed below in Figure 1.

4. Methods

We now proceed to summarize the methods. Many of the details are left for an annex that is available only online as Supplementary Material.

4.1. Identifying clusters

The first step in our methodology is a workable definition to identify inventive clusters as the main geographical unit of our analysis. The geography literature, taking the firm as the unit of analysis, remains close to Porter's (2000, 16) definition of a cluster as '... a geographically proximate group of interconnected companies and associated institutions in a particular field, linked by commonalities and complementarities'. Because we focus on patented inventions instead of firms, the notions of interconnectedness, commonalities and complementarities are substituted by the intensity of knowledge flows, proxied by the number of citations. We define a cluster as a set of geographically close counties among which knowledge flows are particularly strong. The emphasis on knowledge flows in our procedure for identifying clusters not only stems from the common notion that knowledge is an input for producing new knowledge, but also from the idea of technological relatedness. Knowledge flows only between related inventions, and hence we also capture the presence of related activities in the cluster. High inventive activity is a necessary but not sufficient condition for strong knowledge flows.

The cluster identification procedure is motivated by a broad analogy to the idea of metropolitan spaces, in which flows of commuter movements are concentrated (in our case, knowledge flows represent the commuter movements). The clusters are groups of spatially contiguous counties with high patenting activity and intensive knowledge flows between them. We look at citations and patents in the period 2001–2006. The citation matrix, which is directed, gives the number of citations between the US counties. We standardize all cells in this matrix as follows: $cs_{ij} = \left(c_{ij}/(p_i p_j)\right)/(c/p^2)$, where c is the number of citations, p the number of patents, the subscripts i and j indicate counties, and the absence of a subscript indicates an aggregation of counties.

The standardization expresses the number of citations between a pair of counties relative to its expected value, if citations were completely random (within the USA). In this context, 'random citations' means that the relative frequency of citations between a pair of counties is equal to the product of the shares of patents in the two counties. In other words, the more patents there are on either side of the citations relationship, the higher the number of expected citations is. A value lower (higher) than 1 for the standardized citations number indicates less (more) citations than expected randomly. We then binarize the matrix by setting all cells that are larger than 1 to 1, and all other cells to 0. We further thin out the number of ones in this matrix by setting to 0 all cells for which the original number of citations was less than 1, and all cells for which the patents for either the row-county or the column-county is less than 1. The latter part of the procedure is intended to ignore all counties that have very little inventive activity.

In this matrix, we check the geography of all cells with a value of 1. If the row- and column counties for such a cell share a border (we use queen contiguity), we put these counties on a preliminary list. We plot the counties on the list on a map, and check for contiguous areas, which are defined as the clusters. There are 30 clusters that appear in this way, of which 16 have two counties, and the largest cluster has 30 counties. We increase the number of clusters to 35 by merging two clusters that are separated by a sea border, and by including a number of single counties that have an exceptionally large number of patents, and high internal knowledge flows. This is described in detail in the online annex available as Supplementary Material.

4.2. Deep citations

The next step in the methods is to associate the clusters, or geographical space in general, to technological trajectories. As technological trajectories are defined as citation chains (this will be explained in more detail below), we need a way to associate these chains to geography. For this, we will use what we call deep citations. A very simple example is where patent A in region 1 is cited by patent B in region 2, which is in turn cited by patent C in region 3. The knowledge flow is then A(1) → B(2) → C(3). In terms of the start- and endpoint of this example path, we see a 'deep' knowledge flow from region 1 to region 3. However, because patents usually cite more than one other patent, we need a way to conceptualize the complex networks that arise in the real world.

We implement the deep citations idea by looking at all patents in the last 3 years of the period that we used to construct the clusters (2004–2006), and chart their 'ancestry' in terms of patents from 1986–1988. We use a calculation that was pioneered by Martinelli and Nomaler (2014), and that is akin to genealogy. It considers cited patents as the 'parents' of the citing patent, and attributes a 'parenthood' share of $1/n$ to each cited patent, where n is the total number of citations made by the 'child'. By multiplying the direct ancestry shares of different generations, we trace the ancestry across generations. For example, in the genealogy of human reproduction, the share of each parent would be 1/2, the share of each grandparent would be 1/4, and the share of each great-grandparent 1/8, etc. For the patent case we continue tracing back generations as long as we have not yet reached a patent in the period 1986–1988. Obviously, this uses both long and short citation paths, including direct citations between the two periods (i.e. paths of length 2, which do exist but are rare). Also, not all patents from 2004 to 2006 have ancestry in 1986–1988, and we simply ignore those patents that do not have this.

4.3. Technological trajectories in sandwich networks

The deep citations represent all citation paths that exist in geographical space, but technological trajectories are selective important pathways from this large set. The idea is that only those deep citation chains that embody the highest amount of knowledge flows are representative of the main trends in 'technology space', i.e. the trajectories.

Our method for identifying the trajectories is based on the methods proposed by Hummon and Doreian (1989), Verspagen (2007) and Liu and Lu (2012). The patent citation network is directed (knowledge flows from the cited to the citing patent), and also a-cyclical (starting at one node of the network, a path can never return to that node). Two classes of nodes (patents) are of particular interest. A start-point is a patent that is cited, but does not cite any patents. An endpoint is a patent in 2004–2006 that cites other patents, but is not cited itself.

For every citation in the network, we calculate the so-called SPNP (Search Path Node Pair) indicator proposed by Hummon and Doreian. SPNP for the citation of patent p in patent q is counted as follows. First, count all patents in the network for which a path to p exists (including p itself). Then count all patents that can be reached from patent q (including q itself). SPNP is the multiplication of these two counts. It is the number of pairs that can be formed by the patents 'upstream' and 'downstream' the citation. Next, we identify for every start-point in the network the path (ultimately leading to an endpoint) that maximizes the multiplication (sum of logs) of the SPNP values along the path. Such a path is called a main path.

While we will look at deep citations (the 'prequel' to trajectories) at the scale of the entire US inventive space, the specific attributes of trajectories (main paths) are hard to summarize for the entire space, or even the entire list of 35 clusters. Therefore, we look at pairs of the clusters. We start with the full set of patents that forms the entire network of deep citation paths between two specific clusters, and find the main paths (trajectories) in this network. The technological trajectories that we find are therefore true geographical trajectories, as they run between two spatial units. For each pair of spatial clusters, we extract the set of citations that connect the patents in period 2004–2006 in the 'to' cluster to patents in 1986–1988 in the 'from' cluster. This large network, which in-between the start- and endpoints of the citation paths contains many patents not invented in either the to- or from-cluster, is what we call the sandwich network for the specific pair of clusters that we are considering.

5. The US inventive landscape

The starting point of our analysis is the characterization of the US patenting geography by spatial clusters, which embody the peaks in the technological landscape. Map 1 displays the number of patents per county in 2001–2006. White areas indicate zero patents, for colored areas the number of patents increases (roughly exponentially) with darkness. The clusters, identified by the procedure explained above, are displayed in Map 2. Table 1 lists the clusters and gives summary statistics.

Comparing Map 2 and Map 1, we see that there is a large correlation between the number of patents in a county and cluster membership: almost all of the dark areas in Map 1 are also marked in Map 2. Note that this correlation is in no way obvious, as the cluster identification is primarily based on citation intensity, not number of patents (in fact, the number of patents is penalized, as the benchmarking of citation intensity divides by it). The number of patents in the table (2001–2006 period) is distributed rather unequally over the 35 clusters. The largest cluster, i.e. the East coast, holds about 26% of the total, the smallest cluster (Nashville) just under 0.1%. The three largest regions account for just over half of the patents, the bottom-10 clusters for just over 4%. Thus, in terms of the sheer numbers, the inventive landscape in the USA is very peaked, with a small number of leading clusters with very high inventive activity, a larger group of followers, and the far majority of counties (outside the clusters) contributing very little.

6. The geography of direct and deep citations

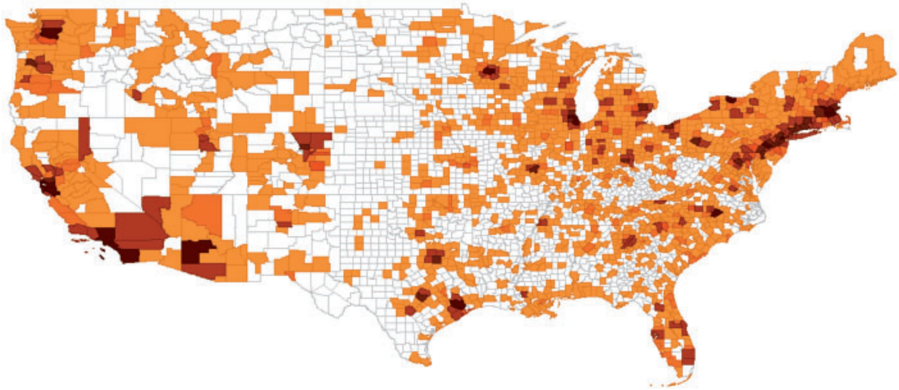
By definition, patent citations are concentrated in the 35 clusters that were identified above. Thus, it is not surprising that direct citations are concentrated within these clusters. However, this is much less obvious for the deep citations (the stuff that trajectories are made of). Because of the indirect linkages that they embody, we may expect that the distribution of the deep citation chains is much less concentrated within single clusters than the direct citations.

As we are interested in comparing distributions of direct and deep citations, we need a benchmark to judge their geographical concentration. Like we did in the procedure that identified the 35 clusters, we construct such a benchmark by referring to the idea of random citations. In this case, it simply means that we break down the total number of cited patents in a specific cluster into the following categories: (i) from within the cluster itself, (ii) from other clusters (between), (iii) from US counties that are not part of a cluster and (iv) non-US countries. The random chance that each of those categories is cited is simply their share in total patents, and this is our benchmark.

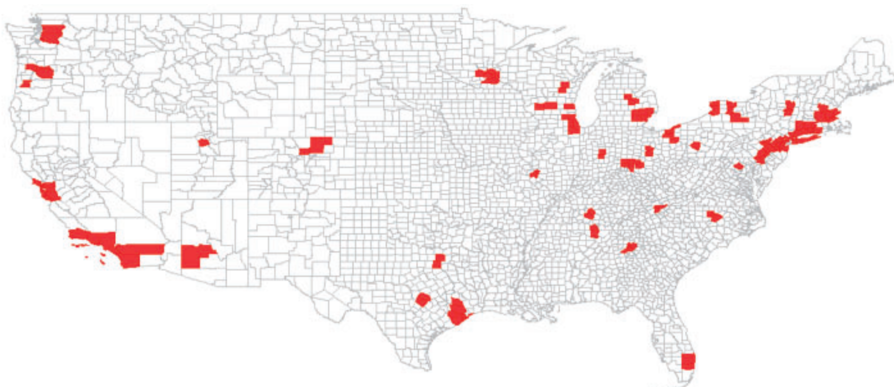
Figure 1 displays the trends in the number of patents for the sub groups of the US cluster regions, the other US counties, and non-US patents. The figure illustrates how the benchmark matters. For example, the number of foreign patents is much larger than that of US patents (the share varies between 67 and 80% over the entire period), and hence the benchmark expects that foreign patents are the largest category in citations made by US patents. Within the USA, the 35 clusters always take the majority of patents: 24% of the total in 1992–1994 and 18% in 2004–2006. In interpreting the trends in Figure 1, remember that the identification of the 35 US clusters was based on patents and citations in the sub period 2001–2006 and that deep citations go back to the period 1986–1988.

Figure 2 compares the distribution of direct citations and deep citations. The figure looks at citations made by (knowledge flowing into) the group of 35 US clusters in the period 2004–2006. The top panel shows the distribution of these citations over the four categories. The sum of each color of the bars is one.

Direct citations in the short run are citations made in 2001–2006, where both the citing and cited patents are from this period. This is the same set of citations that were used to identify the clusters and hence we know that they are strongly spatially



Map 1. Number of patents per county, 2001–2003 and 2004–2006 (white areas indicate zero patents, for colored areas number of patents increases with darkness).



Map 2. Thirty-five clusters, 2001–2003 and 2004–2006.

concentrated. The category of all direct citations includes the entire period 1986–2006, and hence also includes citations over longer time spans than the first category. We see in the top panel that all types of citations are clearly dominated by foreign sources, followed by US clusters (between), and the cluster itself (within). Other US regions are the smallest category.

The bottom panel of the figure invokes the benchmark, by subtracting from the values in the top panel the expected share of citations.[2] This clearly brings out the differences between the types of citations. Direct short-run citations are very much biased to the own cluster (within), much less so but still positively towards other US clusters (between) and US non-cluster regions, and negatively biased towards foreign patents. If we include longer-run direct citations, the bias for the own cluster remains but is weaker, while that of other US clusters increases slightly, and the bias to foreign

2 For deep citations, we also use the benchmark constructed on the basis of citations in the 1986–1988 period.

Table 1. Clusters and summary statistics

Main city/geographical name	Number of counties	Number of patents
East coast (Boston, New York City & Philadelphia)	34	25,348
San Francisco	7	13,862
Los Angeles & San Diego	6	11,635
Phoenix	1	5649
Minneapolis	7	4331
Chicago	5	3547
Seattle	2	3534
Houston	5	2485
Detroit	5	2394
Durham/Chapel Hill	3	2332
Cincinnati	6	2010
Rochester	4	1743
Cleveland	5	1583
Dallas	2	1268
Atlanta	3	1250
Portland	3	1137
Austin	2	998
Montgomery/Washington	1	991
Pittsburgh	1	959
Madison, Wisconsin	2	930
Fort Lauderdale & Palm Beach	2	877
Boulder	2	845
Buffalo	2	815
Indianapolis	2	796
St. Louis	2	739
Benton	1	643
Kingsport	2	597
Salt Lake City	1	494
Schenectady &Albany	3	490
Milwaukee	2	451
Columbus	2	443
Appleton	2	390
Saginaw/Midland	2	356
Huntsville	2	226
Nashville	2	99

patents become much less negative. The deep citations re-enforce this trend. Here the bias to the own cluster is still positive, but smallest among the three citation types. The bias towards foreign countries is still negative, but the absolute value is again smallest among the three types. However, the bias towards other US clusters (between) is largest and positive.

In conclusion, looking at deep citations, we find that these have a broader geographical spread than direct knowledge flows. Although deep citations are still biased to within-cluster flows, they are less so than direct citations. On the other hand, deep citations are also biased, and in a stronger way than direct citations towards flows between clusters. The technological trajectories that we are after are a subset of the deep citation chains analyzed in this section, thus we conclude that the argument of strongly

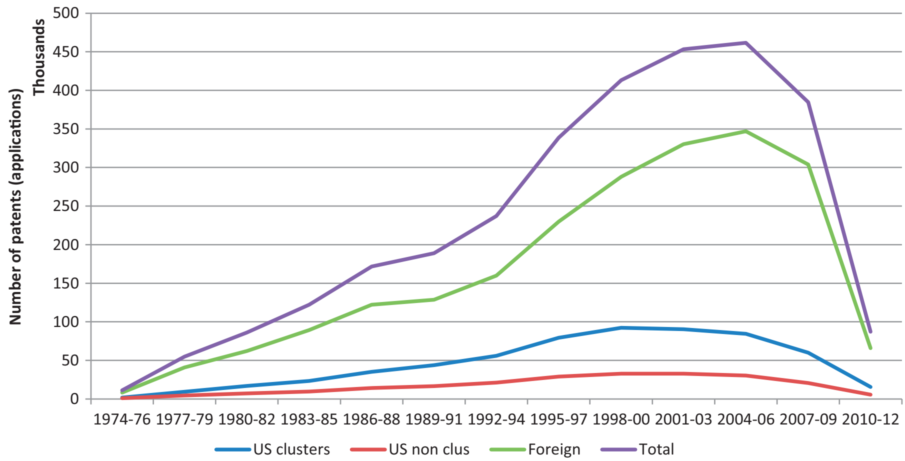


Figure 1. Number of patents in the database.

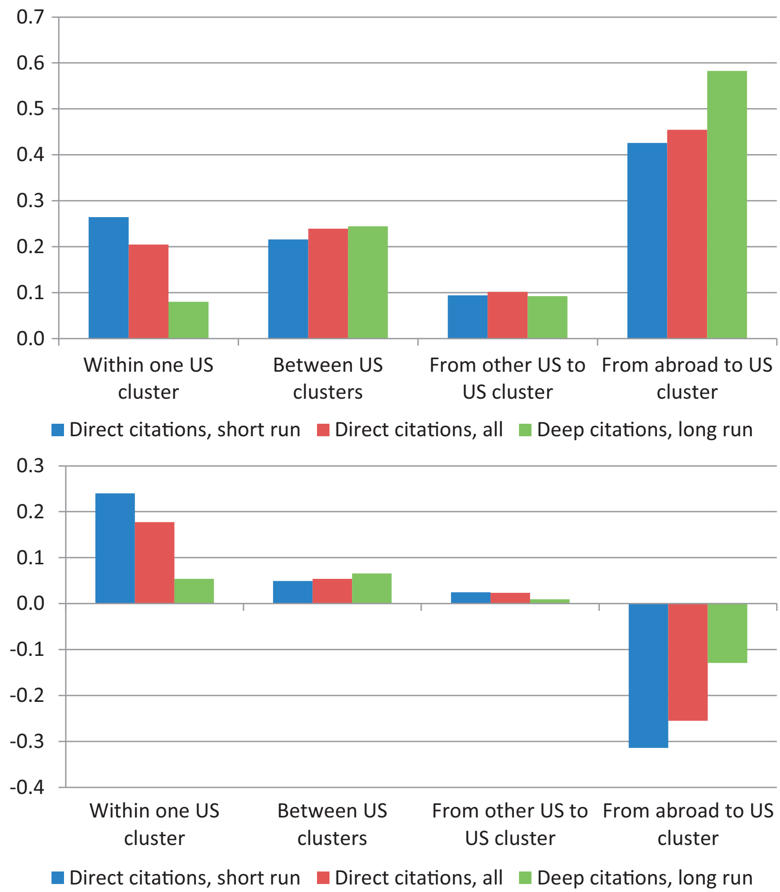


Figure 2. The distribution of direct and deep citations (top panel gives share of all citations, bottom panel is relative to the benchmark).

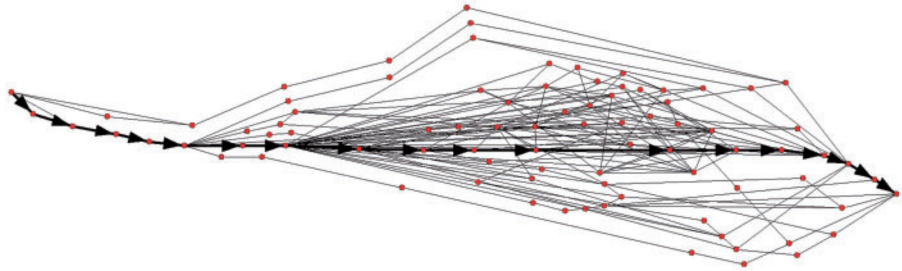


Figure 3. Top main path in the East coast-to-East coast sandwich network (nodes are patents, lines are citations; main path is indicated with arrows and bold lines).

localized knowledge spillovers seems much less relevant for the case of technological trajectories than it is for the incremental changes associated with direct citations.

7. Technological trajectories

We now make the final step of our analysis and extract the actual technological trajectories from the entire set of deep citation chains. Because the category of flows between and within US clusters stands out in magnitude (Figure 2), we focus on this particular category. Results so far only consider the start- and endpoints of the deep citation chains. We now also look at the intermediate nodes on the deep citation chains, to be able to assess which are the main trajectories that cumulative knowledge development takes, and how these trajectories unfold in geographical space.

Figure 3 shows an example of a main path (trajectory), in this case the one that was identified within the East coast-to-East coast sandwich as the main path with maximum multiplicative SPNP. The figure illustrates how the complexity of the underlying network of patent citations can be reduced to a main path or trajectory that highlights the main trends in technology space. The network in the figure is the network of all patents (and citations) in the sandwich that links the start- and endpoint of this main path. Many of the patents in this network, including those on the main path, are patents from a different geographical location than the East coast cluster itself. The main path itself, which consists of 19 patents, contains a 50% share of European inventors, a 30% share of inventors from the 35 US cluster regions, and a 10% share of inventors from US non-cluster regions.

The particular main path in the figure is a trajectory in pharmaceuticals. It starts with four patents on drugs against hypertension and heart disease, then three patents that broaden the range of diseases (including HIV, depression, migraine and psychosis), and finishes with a range of patents on anti-cancer drugs, in particular protein kinase inhibitors. The trajectory has many different firms, including big pharmaceutical firms such as DuPont, Merck and Glaxo, specific gentech firms (AmGen), and smaller pharmaceutical firms (such as the Danish Neurosearch).

The sandwich networks are usually large networks, and hence we cannot look at all 1225 pairs formed by the 35 clusters. Instead, as a sort of case study, we focus on the 16 pairs between the largest four clusters: East coast, San Francisco, Los Angeles/San Diego and Seattle. The 16 sandwich networks together comprise 23,060 main paths.

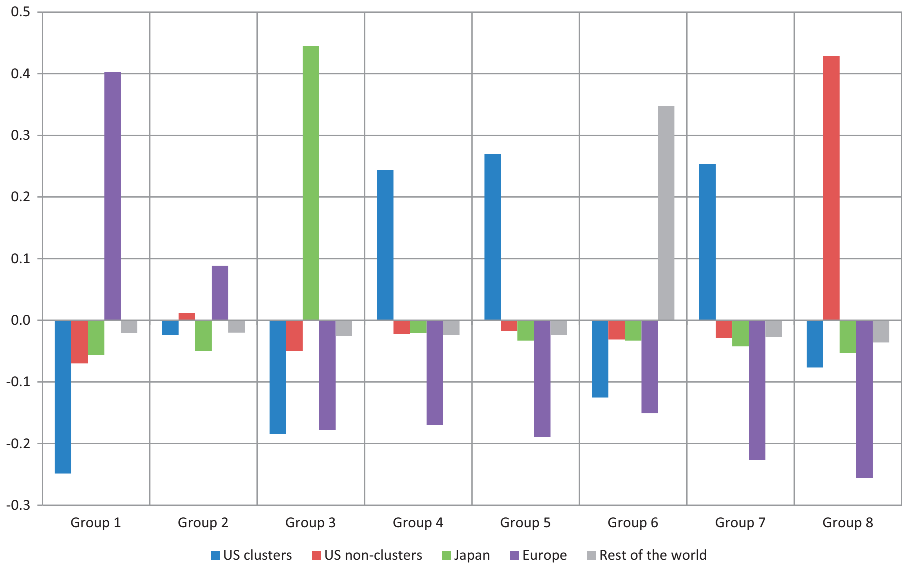


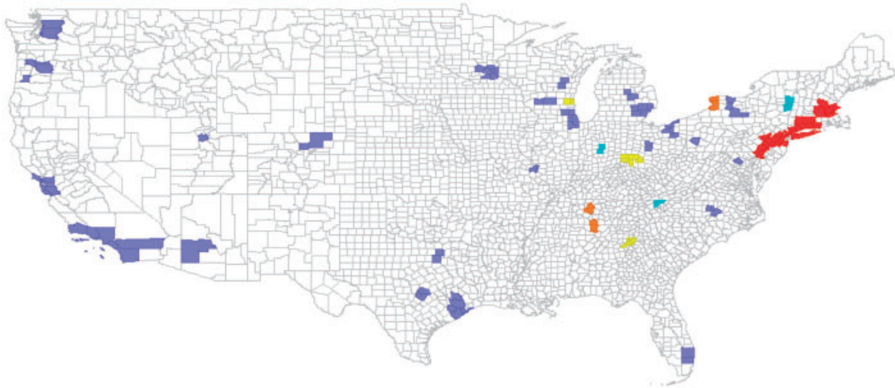
Figure 4. Group characteristics in the cluster analysis (deviations from total sample means).

We characterize these main paths by the share of patents from each of a set of 39 geographical entities (35 US clusters; US non-cluster regions; Europe; Japan; and rest of the world). The 35 US clusters together are the largest category, with a share slightly above one third. Europe follows in second place, with a share just under one third. The other three categories are markedly smaller, with US non-cluster counties as the largest (11%) of these three small categories.

In order to interpret the variety in the composition of the main paths in the sandwich networks, we undertake a cluster analysis, which classifies the 23,060 main paths into a small amount of groups. We use k-means cluster analysis, and settle for a division into eight groups. Choosing eight groups ($k = 8$) is somewhat arbitrary, but we considered alternative groups ($k = 2.10$). Less than eight groups forces some of the interesting between-group heterogeneity that will be discussed below into a single group, while $k = 9$ of 10 does not add qualitatively different insights on the role of US cluster regions, which will be the main focus of our discussion. We have one large group (about one-third of all main paths), which is group 2. Together with the next two largest groups (1 and 4), this group covers almost two thirds of all main paths. The remaining 5 groups vary between 9% and 5% of the total.

Figure 4 characterizes the groups in terms of their composition. Remember that the full detail of the 35 US clusters, plus the 4 other categories, was used to classify the main paths (i.e. to form the groups). Figure 4 displays the mean scores in each of the five categories in the group, with the sample average subtracted. Hence a positive value indicates specialization into this particular category.

Group 2 is the one that has the most homogenous distribution over the five categories. It is most clearly specialized in European contributions to the main paths, but much less so than group 1. It is also slightly specialized in US non-cluster regions, but much less so than group 8. The other groups, including groups 1 and 8, are specialized in only a single category. In group 1, which is one of the larger groups, this is



Map 3. Specialization profile of group 4.

Europe. In group 3, a small group, we have dominance of Japanese patents, in group 6 (also small) dominance of patents from the rest of the world. Groups 4, 5 and 7 are dominated by US cluster regions. Of this, group 4 is relatively large, the other two are small. Group 8 is dominated by US non-cluster counties.

The three groups that are specialized in US cluster regions (4, 5 and 7) can further be described by looking at which clusters play a role. We define an indicator that is akin to the revealed comparative advantage indicator to investigate this. It is defined as $\boldsymbol{R_{ij}} = 100 \times \boldsymbol{S_{ij}}/\boldsymbol{S_i}$, where i (1−35) indicates a cluster region, j (4, 5 or 7) is a group from the cluster analysis, $\boldsymbol{S_{ij}}$ is the share of cluster i in the total contributions of the 35 clusters to the main paths in the group j, and $\boldsymbol{S_i}$ is the share of cluster i in the contribution of the 35 clusters to all (23,060) main paths. A value above (below) 100 indicates a relatively high (low) contribution of the cluster to the group.

Map 3 displays the profile (R values) of group 4, which is the largest of the three US cluster-based groups. The blue colour in the map indicates clusters that have a strongly below average contribution to this group ($R < 75$), cyan indicates $75 < R < 100$, i.e. a mildly below average contribution. The other colours indicate $R > 100$, with yellow $100 < R < 150$, orange $150 < R < 200$ and red $R > 200$.

The map shows that group 4 is strongly influenced by the large East coast cluster. It is the only red cluster on the map. There are only 6 other clusters with $R > 100$ in this group, based in the East side of the country. The entire West coast and most of the Midwest and South have $R < 100$. We can conclude that this is a rather specialized group, in which main paths are strongly concentrated in relatively few clusters. Besides the East coast clusters, the clusters in this group with $R > 100$ are small clusters, i.e. they are low on the list in Table 1. The yellow clusters are Durham/Chapel Hill (ranked 10 in the table), Atlanta (rank 15) and Milwaukee (rank 30). The orange clusters are Buffalo (rank 23), Huntsville (rank 34) and Nashville (rank 35). It seems that these smaller clusters are strongly dependent on the large East coast cluster in terms of being present on the main paths that connect our large clusters in the US inventive system.

Similar maps for groups 5 and 7 confirm the impression from Map 3. Thus, the evidence suggests that these main paths themselves are clustered, instead of consisting of evenly distributed geographical patterns. The main trajectories of knowledge development are like selective rivers unfolding in space, rather than like the wind that blows broadly in a wide circle. The clusters of the US inventive landscape each play a particular role in these deep rivers of main technological trends.

8. Discussion and conclusions

We looked at the geography of technological change, taking a technology perspective rather than the usual perspective of organizations (firms). We use patents and patent citations as the indicator of technological change. Previous findings in the geographical literature have focused primarily on the strong geographical concentration of patent citations (Jaffe et al., 1993), which is in line with the idea that innovation takes place mostly in spatial clusters (cf. Malmberg and Maskell, 2002). Our analysis confirms this strong geographical concentration of patent citations, but in our technological perspective, we interpret it as applying mainly to incremental steps in technological space.

The more substantial part of our analysis looked at accumulated sequences of these incremental direct citations. We call these long-run citation chains deep citations, and we find that these deep citations are especially strong between US incentive clusters, even though within-cluster flows are also important in deep citations (but less so than between-cluster flows). The deep citation chains, by their accumulation of incremental steps, also cover longer technological distances, i.e. they represent cumulative change along which technology takes big steps in the long run. We use methods from a separate literature that uses network theory to map the major trajectories in technological space as the citation chains that attract most knowledge flows in the network of direct citations. By identifying those deep citation chains—the main paths—that capture the largest flows of knowledge, we are able to focus on the particular trajectories of knowledge development that embody the strongest long-run forces of technological change.

Our main finding is that technological trajectories develop between innovation clusters rather than exclusively within clusters. This is akin to the idea that firms in clusters construct and manage global pipelines for knowledge transfer in addition to the local buzz that they find in their own cluster (Bathelt et al., 2004). But contrary to the global pipelines, technological trajectories are a result of collective efforts of firms (and other organizations), and they develop in an evolutionary way as accumulated local change, without a top-down design process. In such an evolutionary process, the forces of change do not likely lead to an outcome that is fully optimal in the sense that economic theory usually assumes. Instead, we may see situations in which firms and other organizations are adapted to local circumstances, including the strategies of their competitors. This may well lead to a situation of lock-in, in which inventors in a particular combination of local clusters jointly focus on a particular technological direction, whereas technological solutions may also be found in other parts of technological space.

The combination between the regional cluster literature and the technological trajectories literature is a novel one. Our results suggest that the geographical dimension is important for the construction of technological trajectories. When a technological trajectory develops, it does so along a specific spatial trajectory. This spatial trajectory mostly consists of chains of patents from inventive clusters. In other words, the results from the technological cluster literature have a strong relevance for the analysis of technological trajectories. The way in which firms use local buzz and build global pipelines will have a strong influence on how trajectories develop. On the other hand, the development of the technological trajectories that the firm is interested in will also determine how it builds its global pipelines of knowledge transfer, and where it will seek local buzz.

This raises questions about firm behavior in the field of technological choice. Which kind of technological resources do firms seek locally, and which ones are sources through global pipelines? Can we even generalize about the answers to these questions, or do the answers differ between geographical locations (clusters)? In general, we suggest that theoretical and empirical work on the geography of innovation takes into account the idea of technological trajectories to develop a more coherent framework for understanding the geographical dimension of technology. For example, the narrative often goes that many of the important trends in ICT come from Silicon Valley, but our results suggest that these trajectories have a much wider geographical base, and are influenced by the global location decisions of firms.

Such a combination of theories of geography and technology may also yield new insights into theorizing about technology itself (e.g. Arthur, 2014). It is likely that technological specialization will play a large role in such a theoretical framework. Because of computational constraints, we have been unable to investigate the role of this factor. However, it is clear that knowledge flows are dependent on technical relations between technology fields, and hence the technological specialization pattern of a cluster will determine from where it can receive the major parts of its technology inflows.

The technological profile of an innovation cluster is, however, not an exogenous factor. Instead, it emerges historically as a result of interaction between local and external actors (inventors, firms, research institutes), and the progress of technological trajectories. Therefore, the technological specialization pattern of inventive clusters will never be the ultimate explanation for the spatial concentration of deep citations. It is a factor that needs to be explained itself, interwoven with the explanation of the concentration of deep citations itself.

Reasoning in the tradition of the innovation clusters literature and the technological trajectories literature would suggest, in our view, that path dependency and lock-in play a role in this process. Innovation clusters depend on interaction (locally and over longer distances) that re-enforces itself by repetition and adaptation. Actors in a local cluster develop routines that become highly specific to their situation. By evolutionary selection, these routines are optimized to become a local fitness maximum, which is specific to the local cluster. Knowledge exchange between local actors and with a selective number of actors outside the local cluster becomes a crucial part of the inputs into the cluster. Obviously, this is a dynamic process, in which change is the norm rather than static equilibrium. Similarly, progress along technological trajectories is cumulative and path dependent. Our analysis suggests that the explanation of regional innovation clusters, their interaction, and of technological trajectories may benefit from more theoretical and empirical linkages between the regional innovation literature and the idea of technological trajectories.

Supplementary material

Supplementary data for this paper are available at *Journal of Economic Geography* online.

References

Arthur, W. B. (2014) *Complexity and the Economy*. Oxford: Oxford University Press.
Bathelt, H. (2005) Geographies of production: growth regimes in spatial perspective 2 – knowledge creation and growth in clusters. *Progress in Human Geography*, 29: 204–216.
Bathelt, H., Malmberg, A., Maskell, P. (2004) Clusters and knowledge: local buzz, global pipelines and the process of knowledge creation. *Progress in Human Geography*, 28: 31–56.
Bathelt, H., Glückler, J. (2011) *The Relational Economy: Geographies of Knowing and Learning*. Oxford: Oxford University Press.
Bottazzi, L., Peri, G. (2003) Innovation and spillovers in regions: evidence from European patent data. *European Economic Review*, 47: 687–710.
Criscuolo, P., Verspagen, B. (2008) Does it matter where patent citations come from? Inventor vs. examiner citations in European patent. *Research Policy*, 37: 1892–1908.
Dicken, P., Lloyd, P. E. (1990) *Location in space: theoretical perspectives in economic geography*, 3rd edn. edn. New York: Harper and Row.
Dosi, G. (1982) Technological paradigms and technological trajectories. *Research Policy*, 11: 147–162.
Jaffe, A. B., Trajtenberg, M., Henderson, R. (1993) Geographic localization of knowledge spillovers as evidenced by patent citations. *The Quarterly Journal of Economics*, 108: 577–598.
Gambardella, A., Harhoff, D., Verspagen, B. (2008) The value of European patents. *European Management Review*, 5: 69–84.
Gertler, M. (2003) Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography*, 3: 75–99
Giura, P., Mariani, M., Brusoni, S., Crespi, G., Francoz, D., Gambardella, A., Garcia-Fontes, W., Geuna, A., Gonzales, R., Harhoff, D., Hoisl, K., Le Bas, C., Luzzi, A., Magazzini, L.Nesta L., Nomaler, Ö., Palomeras, N., Patel, P., Romanelli, M., Verspagen, B. (2007) Inventors and invention processes in Europe: results from the PatVal-EU survey. *Research Policy*, 36: 1107–1127.
Griliches, Z. (1990) Patent statistics as economic indicators: a survey. *Journal of Economic Literature*, 28: 1661–1707.
Hummon, N. P., Doreian, P. (1989) Connectivity in a citation network: the development of DNA theory. *Social Networks*, 11: 39–63
Lau, A. K.W., Lo, W. (2015) Regional innovation system, absorptive capacity and innovation performance: an empirical study. *Technological Forecasting and Social Change*, 92: 99–114.
Liu, J. S., Lu, L. Y. Y. (2012) An integrated approach for main path analysis: development of the Hirsch index as an example. *Journal of the American Society for Information Science and Technology*, 63:528–542.
Malmberg, A., Maskell, P. (2002) The elusive concept of localization economies: towards a knowledge-based theory of spatial clustering. *Environment and Planning A*, 34: 429–449.
Mina, A., Ramlogan, R., Tampubolon, G., Metcalfe, J. S. (2007) Mapping evolutionary trajectories: applications to the growth and transformation of medical knowledge. *Research Policy*, 36: 789–806.
Martinelli, A., Nomaler, Ö. (2014) Measuring knowledge persistence: a genetic approach to patent citation networks. *Journal of Evolutionary Economics*, 24: 623–652.
Marshall, A. (1890), *Principles of Economics*. London: MacMillan.
Maurseth, P.-B., Verspagen, B. (2002) Knowledge spillovers in Europe. a patent citations analysis. *Scandinavian Journal of Economics*, 104: 531–545.
Morgan, K. (2004) The exaggerated death of geography: localized learning, innovation and uneven development. *Journal of Economic Geography*, 4: 3–21.
Nuvolari, A. Verspagen, B. (2009) Technical choice, innovation, and British steam engineering, 1800–50. *Economic History Review*, 62: 685–710.
Owen-Smith, J., Powell, W. W. (2004) Knowledge networks as channels and conduits: the effects of spillovers in the Boston biotechnology community. *Organization Science*, 15: 5–21.
Porter, M. (2000) Location, competition, and economic development: local clusters in a global economy. *Economic Development Quarterly*, 14: 15–34.
Sahal, D. (1981) *Patterns of Technological Innovation*. Boston: Addison-Wesley.

Saxenian, A. (1994) *Regional advantage, culture and competition in Silicon Valley and Route 128*. Cambridge, MA: Harvard University Press.

Sorenson, O. (2003) Social networks and industrial geography. *Journal of Evolutionary Economics*, 13: 513–527.

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

Storper, M. (1993) Regional worlds of production: learning and innovation in the technology districts of France, Italy and the USA. *Regional Studies*, 27: 433–455.

Thompson, P. (2006) Patent citations and the geography of knowledge spillovers: evidence from inventor – and examiner – added citations. *Review of Economics and Statistics*, 88: 383–388.

Verspagen, B. (2007) Mapping technological trajectories as patent citation networks: a study on the history of fuel cell research. *Advances in Complex Systems*, 10: 93–115.