

Papers in Evolutionary Economic Geography

13.04

Scientific Knowledge Dynamics and Relatedness in Bio-Tech Cities

Ron Boschma, Gaston Heimeriks, Pierre-Alexandre Balland



Utrecht University
Urban & Regional research centre Utrecht

Scientific Knowledge Dynamics and Relatedness in Bio-Tech Cities

Ron Boschma *

Gaston Heimeriks ^β

Pierre-Alexandre Balland [†]

Abstract

This paper investigates the impact of scientific relatedness on knowledge dynamics in biotech at the city level during the period 1989-2008. We assess the extent to which the emergence of new research topics and the disappearance of existing topics in cities are dependent on their degree of scientific relatedness with existing topics in those cities. We make use of the rise and fall of title words in scientific publications in biotech to identify major cognitive developments within the field. We determined the degree of relatedness between 1,028 scientific topics in biotech by means of co-occurrence of pairs of topics in journal articles. We combined this relatedness indicator between topics in biotech with the scientific portfolio of cities (i.e. the topics on which they published previously) to determine how cognitively close a potentially new topic (or an existing topic) is to the scientific portfolio of a city. We analyzed knowledge dynamics at the city level by looking at the entry and exit of topics in the scientific portfolio of 276 cities in the world. We found strong and robust evidence that new scientific topics in biotech tend to emerge systematically in cities where scientifically related topics already exist, while existing scientific topics had a higher probability to disappear from a city when these were weakly related to the scientific portfolio of the city.

Key words: relatedness, co-occurrence analysis, knowledge dynamics, geography of biotech, title words

JEL codes: O33, R11, L65, D83

* CIRCLE, Lund University, and Department of Economic Geography, Urban & Regional research centre Utrecht (URU), Utrecht University

Ron.Boschma@circle.lu.se

^β Department of Innovation Studies, Copernicus Institute, Utrecht University

g.j.heimeriks@uu.nl

[†] Department of Economic Geography, Urban & Regional research centre Utrecht (URU), Utrecht University Eindhoven Centre for Innovation Studies (ECIS), Eindhoven University of Technology

p.balland@uu.nl

1. Introduction

The geography of scientific knowledge production is very uneven. For instance, the world's most influential scientific researchers reside in a very small number of cities (UNESCO, 2010). This is reinforced by research linkages which connect in particular the scientific hubs (Gertler and Levitte, 2005; Laudel, 2005; Gittelman, 2007; Zucker and Darby, 2007; Hoekman et al., 2009; Tripl, 2009). At the same time, there is a process of ongoing globalization in scientific research (Cooke, 2006; Moodysson, 2007; Hoekman, 2012), as illustrated by the ever increasing number of countries that contribute to scientific publications. Scientific knowledge production in biotech, for instance, has shifted away in relative terms from the US towards Asian regions, like Seoul, Tokyo, Beijing and Singapore, which have become world players in biotech (Heimeriks and Boschma, 2013).

This calls in question how knowledge in biotech science evolves over time, and to what extent regions still contribute to this evolution in the context of globalization. Where is new scientific knowledge created, and to what extent does new knowledge build on existing regional knowledge? Recent studies show that product relatedness is a main driver of industry dynamics at the regional scale, as new industries tend to build on and exploit capabilities in related regional industries, and existing industries are more likely to disappear when few or no related industries are present in the region (Neffke et al., 2011; Boschma et al., 2013). The question is whether the rise and fall of scientific knowledge is also depending on the degree of relatedness with existing knowledge, and whether the body of scientific knowledge available at the regional level matters in that respect. Systematic evidence is yet lacking.

The main objective of the paper is to investigate the impact of scientific relatedness on knowledge dynamics in biotech science at the city level worldwide during the period 1989-2008. More in particular, we assess the extent to which the emergence of new scientific topics in biotech and the disappearance of existing topics in cities are dependent on their degree of relatedness with existing topics present in those cities. In order to measure knowledge dynamics in biotech science, we make use of title words in scientific publications over a long period of time, in order to identify the rise and fall of key scientific topics in biotech. Inspired by the 'product space' concept (Hidalgo et al., 2007), we construct a 'scientific space' in which the degree of relatedness between 1,028 scientific topics in biotech is determined by means of co-occurrence analysis. Then, we combine this relatedness indicator between topics in biotech with the topic portfolio of 276 biotech cities worldwide, in order to determine how

close a new topic that entered a city (and an existing topic that exited a city) is to the scientific portfolio of that city. We analyzed knowledge dynamics at the city level by looking at the entry and exit of topics in the scientific portfolio of 276 cities in the world. Our main finding is that new scientific topics in biotech emerge systematically in cities where scientifically related topics already exist, while existing topics were more likely to disappear from a city when these were weakly related to the scientific portfolio of the city.

The paper is structured as follows. Section 2 sets out the main theoretical ideas, especially the process of branching of scientific knowledge dynamics at the regional level. Section 3 describes the methodology and the data used. We explain how we define the relatedness between scientific topics in biotech, and how we use that information as an input for our econometric exercise in which we assess the impact of scientific relatedness on the rise and fall of topics in biotech science. Section 4 presents the main findings. Section 5 concludes.

2. Scientific knowledge dynamics, relatedness and regional branching

In evolutionary thinking, knowledge production is often described as a cumulative, interactive and path-dependent process (Dosi, 1982; Nelson and Winter, 1982). Because of bounded rationality, search for new knowledge is highly uncertain. As a result, agents tend to draw on knowledge acquired in the past, which provides opportunities but also sets limits to what can be learned (Atkinson and Stiglitz, 1969; Heiner, 1983). As Cohen and Levinthal (1990) argued, agents are more likely to understand, absorb and implement external knowledge when it is close to their own knowledge base. This also implies that knowledge is widely dispersed among many heterogeneous agents, and that the process of knowledge creation heavily depends on combining different capabilities of agents (Antonelli, 1995; Nooteboom, 2000). Therefore, interaction between agents is central, and knowledge production is more than ever an outcome of socially constructed learning processes (Amin and Cohendet, 2000).

This path-dependent and interactive nature of knowledge production becomes manifest not only at the organizational but also at the regional level. To transfer knowledge in an effective manner, not only absorptive capacity of actors but also close and intensive face-to-face contacts between actors are needed. Geographical proximity accommodates this type of interaction, especially for more tacit forms of knowledge (Gertler, 2003). So, knowledge dynamics unfolds at the level of organizations but is situated in a social and geographical context. As regions specialize in particular competences, these offer opportunities to local

organizations for further improvements in similar fields of knowledge, and discourage the creation of knowledge that does not match the regional knowledge base (Boschma, 2004). In sum, the regional accumulation of tacit knowledge provides an intangible asset that is difficult to cope by non-local agents, as geographical distance may form an insurmountable barrier for the transfer of tacit knowledge (Maskell and Malmberg, 1999).

Knowledge also accumulates at the regional level because some mechanisms through which knowledge diffuses across organizations, like spinoff activity, labor mobility and social networking, are often spatially bounded (Capello, 1999; Boschma and Frenken, 2011). That is, new spinoff companies tend to locate in the same region as their parent organization, where these new ventures exploit the knowledge they acquired from the parent (Klepper, 2007). And most employees still change jobs in the same labor market region, which means that the transfer of knowledge and skills through labor mobility primarily occurs between local organizations (Eriksson, 2011). And there is evidence that knowledge sharing through social networks tends to be often local (Breschi and Lissoni, 2009).

This is further reinforced by the regional institutional context which shapes the interaction between agents. As knowledge is more and more distributed among heterogeneous actors, there is a strong need to connect, combine and integrate different capabilities. Institutions (like cultural values, conventions and social practices) are important enablers to deal with this increasing complexity of knowledge creation, because they decrease uncertainty and create mutual understanding between actors, especially at the regional level, as geographical proximity favors institution building (Storper, 1995). Accordingly, linkages among agents across different institutional contexts tend to be less widespread and are not always very productive in terms of learning and innovation (Gertler, 2003). Moreover, these regional institutional settings are difficult to replicate in other places as these have evolved over long periods of time. In other words, knowledge often accumulates at the regional level, in which the territory shapes combinatorial knowledge dynamics by providing physical proximity and institutional closeness between agents (Strambach and Klement, 2012).

This is not to say that regions only drive knowledge dynamics. On the contrary, there are tendencies of globalization in knowledge production, and it is widely accepted that knowledge dynamics is a multi-scalar phenomenon (Asheim and Isaksen, 2002; Bathelt et al., 2004; Moodysson, 2007; Martin, 2012). Nevertheless, place still matters in processes of collective interactive learning. Studies have recently demonstrated that region-specific capabilities

operate as sources of diversification. That is, regions are more likely to expand and diversify into sectors that are closely related to their existing activities (Neffke et al., 2011; Boschma et al., 2013). This means that geographically localized capabilities provide opportunities but also set constraints for regions to diversify into new industries (Neffke, 2009). Boschma and Frenken (2011) describe this process as regional branching, in which new industries arise from technologically related industries in regions in which existing competences are recombined. Moreover, apart from the fact that sectors that are technologically related to other sectors in the region are more likely to enter, Neffke et al. (2011) also found that sectors have a higher probability to exit a region when these have few or no other sectors in the region to which they are technologically related. In other words, the rise and fall of industries is heavily conditioned by the presence of technologically related industries in the region.

Although the literature on the evolutionary geography of scientific knowledge production is still underdeveloped (Frenken et al., 2009; Hoekman, 2012), it seems rather straightforward to apply this line of reasoning to scientific knowledge dynamics. Especially in science, knowledge is widely dispersed among many heterogeneous agents, and it is increasingly difficult for a researcher to possess the necessary skills and knowledge to solve scientific problems alone (Cronin et al., 1998; Wuchty et al., 2007; Hardeman, 2012). Therefore, the spatial concentration of research may bring all kinds of advantages like lower search costs for research partners and new personnel (Carvalho and Batty, 2006). In their programmatic paper on spatial scientometrics, Frenken et al. (2009) explained why interactions in science tend to be spatially biased towards physically proximate actors. There is ample evidence that research collaboration is indeed triggered by geographically and socially proximate partners (Breschi and Lissoni, 2009), as in the case of university-industry research collaboration (Katz, 1994; Hoekman et al., 2009). Scientific knowledge production is also heavily affected by territorial institutions such as funding schemes, labour regulations and property right regimes, which constrain mobility to a considerable degree, especially across national institutional boundaries. But even when there is international mobility of researchers, it tends to reproduce the spatial concentration of science to a considerable degree, as leading scientists tend to be attracted to the major scientific hubs (Laudel, 2005; Zucker and Darby, 2007; Tripl, 2009).

This does not mean that scientific knowledge production is a completely geographically localized phenomenon. Globalization of knowledge production in science is occurring, and it is heavily promoted by the harmonization of national institutional frameworks and the set-up of international research programs. Moreover, science is a global, collective and distributed

system, a collection of individuals and organizations contributing to a common body of scientific knowledge that is shared at conferences and workshops and codified in journals and reports (Wagner, 2008; Heimeriks, 2012). This global body of scientific codified knowledge acts as a focusing device for the whole scientific community. Kauffman (1995) has referred to this as the 'adjacent possible', which defines all new knowledge that can be directly derived from the present state of knowledge. As such, it provides a kind of map of possible new scientific topics that may occur from the current set of topics.

In this context, it then becomes a crucial question whether the emergence of new scientific topics is triggered and channeled by the adjacent possible at the local scale. In the literature on the geography of scientific knowledge production, little attention has yet been devoted to this question. The knowledge base literature (Asheim and Coenen, 2005; Asheim and Gertler, 2005) claims that knowledge production and learning processes are less sensitive to geographical proximity in sectors like biotech which are grounded in analytical knowledge bases. Studies have shown that knowledge relationships in analytical knowledge bases indeed take place over larger geographical distances, as compared to other knowledge bases (Martin and Moodysson, 2012). However, one still wonders what the role of the region is in scientific knowledge production. As Hoekman et al. (2010) have demonstrated, physical co-presence remains important for exchanging complex forms of scientific knowledge. Therefore, it is crucial to know whether there still is a systematic tendency of new scientific knowledge to be strongly rooted in capabilities and institutional practices at the local scale, despite globalization.

Heimeriks and Boschma (2013) explored whether new and disappearing topics in different types of cities were related to the existing set of topics that is available worldwide. Taking the global set of research topics as reference, they provided evidence for the process of branching and the adjacent possible at the global scale, and how that works out in different city contexts. Doing so, they described the evolution of scientific knowledge as a dynamic interplay between the global and the local scale, in which codified and accessible knowledge available in leading international journals interacts with geographically localized processes of scientific knowledge production. However, they did not investigate the extent to which the emergence of new scientific topics and the disappearance of existing topics in cities are dependent on their degree of relatedness with existing topics available in those cities. Such a study would provide evidence for the branching process in scientific knowledge at the city scale.. It would shed light on the cumulative and path dependent nature of scientific knowledge production at

the regional scale, and the extent to which this is driven by region-specific capabilities. We argue that the scope of opportunities for researchers to contribute within the constraints of the existing body of scientific knowledge may be very different across locations, and what constitutes a research opportunity and how it is dealt with, may be locally situated.

3. Data and methodology

The main argument developed in this paper is that the dynamics of scientific knowledge is a path and place dependent process, and that the scientific composition of cities influence their further capacity to evolve. More particularly, we aim to evaluate the impact of scientific relatedness on the emergence of new topics in science at the city level. Our methodology follows the "product space" framework, which integrates network science to macroeconomic theories in order to understand the uneven development of countries (Hausmann and Klinger, 2007; Hidalgo et al., 2007). This framework develops a 2-mode network approach of the economy constructed from country-product pairs (Hidalgo et al., 2007). In this paper, we apply the product space framework to scientific knowledge dynamics, and our 2-mode network is based on pairs of city-topics constructed from publication records in the biotech field from 1986 to 2008.

Data and context

We use the accumulated body of codified knowledge in biotech for the period 1986-2008 as laid down in scientific publications. The field of biotech is delineated using journal-journal citation patterns (Leydesdorff and Cozens, 1993). This method is based on factor analysis of the journal-journal citations matrix of the central journals in biotech, which is *Biotechnology and Bioengineering*. The relational citation environment of that journal can be determined by using a threshold of 1%. We include all journals in the analysis that are cited by the core journal at least 1/100th of the total number of citations. For the resulting set of journals, a journal-journal citation matrix is constructed, with citing behavior as variables. A factor analysis of this matrix resulted in factors consisting of journals that revealed similar citation patterns. The factor on which the core journal had it highest loading represents the field under study. The other factors represent a set of research fields that are related to the field. In recent years, it seems that there is a narrow definition of the field consisting of only three core journals that we used for our analyses: *Biotechnology Progress*, *Biotechnology and*

Bioengineering, and *Journal of Biotechnology*. These three journals are expected to provide an accurate representation of developments in biotechnology, because they are consistently present in the core of the field in the period under study (Leydesdorff and Heimeriks 2001, Heimeriks and Leydesdorff 2012). After obtaining this set of journals, all publications for the period 1986-2008 were downloaded from the Web of Science. The total set of publications consisted of 13,386 articles.

The use of title words in the set of biotech publications provides us with an indication of the cognitive developments of topics within the field¹. Our publication records provide title words, in addition to institutional addresses of universities and companies authors are affiliated to. Several indicators based on title words have been developed to trace the development of science (e.g. Leydesdorff, 1989). Scientific texts carry information indicated by selections of title words and co-occurrences of title words (Callon et al., 1986; Leydesdorff, 1991). Title words are carefully selected by authors in order to position their knowledge claims. As such, “title words seem to offer a means of making visible the internal cognitive structure” (p. 217) of a field (Leydesdorff 1989).. The information is positioned in a network with an emerging (and continuously reconstructed) structure (Leydesdorff, 2010). In this way, an evolving discourse of scientific topics can be measured by using title words and their co-occurrences as the observable variation. For the purpose of our study, plural morpheme stemming was applied (e.g. removing the "-s" at the end of "computers").

The scientific relatedness between topics

To measure the relatedness between a topic and the scientific profile of a city, we first have to measure the scientific relatedness between topics. Instead of using journal's classifications, we follow Hausmann and Klinger (2007) and compute an outcome-based measure of relatedness based on co-occurrence analysis. Previous studies have used co-occurrences of products within countries (Hidalgo et al., 2007), products within regions (Boschma et al., 2013), products within plants (Neffke et al., 2011), or co-occurrences of technology classes within patents (Boschma et al., 2012). In this paper, we use a simple and normalized measure of relatedness based on the co-occurrences of fine-grained scientific topics, i.e. title words, within journal articles. The literature presents many meaningful examples of topic mappings

¹ Abstract words and keywords could also be used for tracing the evolutionary development of topics in a field. They could not be used in this study covering the period 1986-2008 however, because abstracts and keywords were included in the ISI Web of Science from 1991 onwards.

using co-occurrences of words (e.g. Lucio-Arias and Leydesdorff 2009, Heimeriks and Leydesdorff 2012 as recent examples) Researchers select words to describe their research topic and within the context of a body of literature the words acquire their specific meaning. In this way we can account for different ways of using words, and for changes over time in the meanings of words (Van den Besselaar and Heimeriks, 2006).

Two topics are considered to be related if they co-occur on many journal articles, because they are assumed to require the same scientific capabilities. To normalize co-occurrences, we use the Jaccard index, which has widely been used in scientometrics². With occ_{ij} denoting the total number of times i and j co-occur in the same journal article, and occ_i denoting the total number of occurrences of i ³, the relatedness $\varphi_{i,j,t}$ between each topic i and j is given by:

$$\varphi_{i,j} = \frac{occ_{ij}}{occ_i + occ_j - occ_{ij}} \quad (1)$$

As a result, the measure is symmetric and $\varphi_{i,j,t} \in [0,1]$. A value of 0 indicates that the two topics never co-occurred within the same journal article, while a value of 1 indicates that the two topics systematically co-occur. Since topics might be related at some point of time but not anymore at a later stage (and the other way round), as the scientific field evolves, we compute the relatedness $\varphi_{i,j,t}$ between each pair of topic i and j for five different non-overlapping periods of time from 1986 to 2008⁴: 1986-1988; 1989-1993; 1994-1998; 1999-2003; 2004-2008. For instance, on the period 2004-2008, the relatedness between "cell" and "protein" is equal to 0,10⁵. Table 1 indicates the relatedness between the 10 most central title-words on this scientific space from 2004 to 2008. As we can see in Table 2, the correlation between these relatedness matrices over time is important given the fine grained classification we use and the substantial rates of entry and exit of title-words.

² Although the Jaccard index is popular in scientometrics, there is still a debate concerning the most appropriate measure for normalization of co-occurrence data. For a discussion of alternative normalization procedures, the reader is referred to Leydesdorff (2008) and Van Eck and Waltman (2009).

³ If i occurs on 5 journal articles in total during a given period, $occ_i = 5$, regardless of the number of words with which i co-occurs. Another alternative would have been to use the total number of times i co-occurs with other words to compute occ_i .

⁴ Excepted the first period for which we only have 3 years of data, all the periods run on five years.

⁵ The title-word "cell" and the title-word "protein" co-occurred 128 times on journal articles from 2004 to 2008, while singly, "cell" occurred 868 times and "protein" occurred 539 times. Therefore $\varphi_{\text{cell, protein, 2004-2008}} = 128 / (539 + 868 - 128) = 0.10$.

Table 1. Scientific relatedness between the 10 most central topics (2004-2008)

	cell	Rec.	cult.	prod.	coli	protein	Esch.	effect	batch	Expr.
cell	1	0.08	0.17	0.1	0.04	0.1	0.03	0.06	0.04	0.08
recombinant	0.08	1	0.05	0.14	0.16	0.16	0.15	0.04	0.05	0.1
culture	0.17	0.05	1	0.11	0.02	0.03	0.02	0.06	0.1	0.03
production	0.1	0.14	0.11	1	0.09	0.1	0.09	0.06	0.06	0.03
coli	0.04	0.16	0.02	0.09	1	0.09	0.86	0.04	0.03	0.09
protein	0.1	0.16	0.03	0.1	0.09	1	0.08	0.03	0.03	0.11
Escherichia	0.03	0.15	0.02	0.09	0.86	0.08	1	0.04	0.03	0.08
effect	0.06	0.04	0.06	0.06	0.04	0.03	0.04	1	0.02	0.02
batch	0.04	0.05	0.1	0.06	0.03	0.03	0.03	0.02	1	0.01
Expression	0.08	0.1	0.03	0.03	0.09	0.11	0.08	0.02	0.01	1

Note: Relatedness is given by the co-occurrences of title-words ($N=1,028$) on journal articles from 2004 to 2008, normalized by a Jaccard index (Eq. 1).

Table 2. The dynamics of scientific relatedness between topics over time

	1986-1988	1989-1993	1994-1998	1999-2003	2004-2008
1986-1988	-	-	-	-	-
1989-1993	0.275***	-	-	-	-
1994-1998	0.273***	0.407***	-	-	-
1999-2003	0.255***	0.376***	0.456***	-	-
2004-2008	0.236***	0.347***	0.420***	0.481***	-

Note: This table is based on the relatedness between each pair of 1,028 topics. The change in scientific relatedness is inflated by the entry and exit of topics over time

The scientific relatedness between topics and cities

We then combine the relatedness $\varphi_{i,j,t}$ between topics with the scientific specialization of cities (the topics on which they publish) to construct a city-topic level variable that indicates how close a potential new topic is to the existing scientific portfolio of a given city. The portfolio of cities is constructed from the addresses mentioned in journal articles.

The publications can be attributed to cities as unit of geographical analysis. Each publication in our dataset contains one or more institutional addresses that enable us to specify the location of the universities, research organisations and companies, to which the authors are affiliated. In this study, publications are fully attributed to each author location (no fractional

counting was applied). All addresses in the publication set could be provided with geo-coordinates at <http://www.gpsvisualizer.com/geocoder/>. Yahoo! was used for obtaining the city names. As such, we can attribute topics to city locations in each year. From this geographical information, we can derive information about local path-dependent dynamics and collaboration patterns based on co-authorships (Leydesdorff & Persson, 2010).

This relatedness variable combines therefore relatedness between topics with the particular scientific expertise of cities and it indicates how cognitively close a potential new topic is to the pre-existing set of topics in a given city is specialized. The variable is constructed in the same way than the "density" index developed by Hidalgo et al. (2007) or the "closeness" index proposed by Neffke et al. (2011). The relatedness of a topic i to the expertise of city c in time t is computed by summing the scientific relatedness⁶ of this topic i to all the scientific topics that can be found in the portfolio of the city c , divided by the sum of scientific relatedness of topic i to all the other topics in t . We then multiply by 100. The city-topic relatedness can then easily be interpreted as the percentage of related topics found in a city. For instance, if a given topic is scientifically related to 100 other topics (on 1,028), and if a city c is specialized in 10 (numerator) of these 100 topics (denominator), then the relatedness between topic i and city c will be equal to $(10/100)*100 = 10\%$. By construction, the relatedness variable lies then between 0 % and 100 % and is given by the following formula:

$$RELATEDNESS_{i,c,t} = \frac{\sum_{j \in c, j \neq i} \varphi_{ij}}{\sum_{j \neq i} \varphi_{ij}} \times 100 \quad (2)$$

Table 3 presents some real measures of relatedness between the top 10 words and the top 10 cities that appeared on scientific articles most frequently during the period 1999-2003. For instance, the degree of relatedness between the city "Cambridge, MA" and the topic "cell" is 67%, which means that researchers from Cambridge, MA are specialized in 67% of the topics that are scientifically related to "cell".

⁶ To further ease the interpretation of the city-topic relatedness index, we first had to dichotomize the topic-topic measure of relatedness. For each period, we use a 5% threshold, which means that only the top 5% (in terms of $\varphi_{i,j}$) of all technology-pairs are considered as related (=1), while the other pairs of topics are considered as unrelated (= 0). The main findings presented in the paper are robust to alternative dichotomization procedures (1% and 10% threshold) and to using the weighted relatedness index.

Table 3. Scientific relatedness between top cities and hot topics

$City_c$	$Topic_i$	$Relatedness_{i,c,t}$
Cambridge (USA)	Cell	67 %
Lund (Sweden)	Production	69 %
Taejon (South Korea)	Protein	58 %
London (UK)	Culture	60 %
Wageningen (Netherlands)	Effect	66 %
Zurich (Switzerland)	Recombinant	62 %
Delft (Netherlands)	Coli	55 %
Seoul (South Korea)	System	54 %
Baltimore (USA)	Expression	49 %
Pittsburgh (USA)	Escherichia	32 %

Note: City-topic relatedness index are computed from relatedness between topics and the scientific expertise of cities from 1999 to 2003.

Sample construction

We generated our sample by identifying all biotech publications through journal-journal citation analysis as described above. After obtaining this set of journals, all publications for the period 1986-2008 were downloaded from the Web of Science. The total set of publications consisted of 13,386 articles involving 1,789 unique cities and 6,430 unique title-words. Since most of the cities and words appear only very few times, we concentrate on topics and cities that are scientifically important to avoid statistical noise. The results presented in the paper concerns words and cities that can be associated with around 1 paper a year⁷. Our sample finally includes 276 cities and 1,028 fine-grained scientific topics for 4 non-overlapping periods (the independent variables are lagged by one period) resulting in a total of 1,134,912 city-topic observations. We analyzed knowledge dynamics at the city level by looking at the entry/exit of topics in the scientific portfolio of cities. From this sample we excluded the observations with topics that were, by construction, not able to emerge in a given city (if the word was already in the scientific portfolio in the previous period). And vice-versa, we excluded topics that were not able to exit cities because they did not belong to the pre-existing set of topics. Table 4 provides some summary statistics and the correlation matrix of the variables used in the econometric analysis.

⁷To be included in the analysis, cities need to publish more than 20 papers on the entire period, while topics need to appear on more than 20 papers.

Table 4. Summary statistics

	Mean	SD	Min	Max	Obs.	[1]	[2]	[3]	[4]
[1] Entry	0.045	0.208	0	1	1,089,104	-	-	-	-
[2] Exit	0.685	0.465	0	1	45,808	-	-	-	-
[3] Scientific relatedness (%)	6.996	9.882	0	100	1,134,912	0.127	-0.109	-	-
[4] # publications (topic level)	15.232	30.741	0	591	4,112	0.127	-0.308	0.231	-
[5] # publications (city level)	10.886	15.57	0	146	1,104	0.090	-0.170	0.486	0.237

Note: to correspond with the econometric analysis, the variables "# publications (topic level)" and "# publications (city level)" have been log-transformed before computation of the correlation matrix.

4. Econometric estimations

Our objective is to estimate empirically the impact of scientific relatedness on knowledge dynamics at the city level. Our unit of analysis is the city-topic observation, and to explain the emergence of new topics in science, we regress the entry of new title-words on their degree of scientific relatedness with the scientific portfolio of cities:

$$P(ENTRY_{i,c,t}) = \beta_0 + \beta_1(RELATEDNES S_{i,c,t-1}) + \beta_2(CITY_{c,t-1}) + \beta_3(TOPIC_{i,t-1}) + \varepsilon_{i,c,t} \quad (3)$$

The dependent variable $ENTRY$ in Eq. (3) is a dummy that takes a value of one when a topic that did not belong to the scientific portfolio of the city c in time $t-1$ enters its portfolio in time t , and zero otherwise. All the variables are centered around the mean, so the intercept β_0 gives the expected mean probability of entry when all the independent variables in the equation are set to their means. The key explanatory variable $RELATEDNES S_{i,c,t-1}$ indicates how related the potential new topic i is to the pre-existing knowledge set of capabilities of c ; But to capture the relationship between scientific relatedness and knowledge dynamics at the city level in science it is important to control for the scientific importance of both topics and cities. In fact, some words tend to emerge in many cities (such as the topic "cell"), while some cities tend to have expertise in a wide variety of topics (such as the city "Cambridge, MA"). A first control variable is constructed at the city level: $CITY_{c,t-1}$, by computing the natural logarithm of the total number of publications of this city on a given 5-years period. In a similar vein, the popularity of topics is controlled by a second variable: $TOPIC_{i,t-1}$ constructed

by computing the natural logarithm of the total number of publications in which a word occur on a given 5-years period. $\varepsilon_{i,c,t}$ is a regression residual. We estimate Eq. (3) by using a linear probability (OLS) regression⁸ and standard errors are clustered at the city and topic level⁹ to correct for correlation within groups of observations (Wooldridge, 2003).

Table 5 presents the results for the estimation of Eq. (3). The baseline model (model 1) simply regresses the entry of new scientific topics on their degree of relatedness with the existing portfolio of cities (lagged by one period). As the variables are centered around their mean, the intercept is equal to 0.04534, which corresponds then by construction to the mean of the entry variable (see table 4). Therefore, the probability that a new topic emerges in a city is around 4.5 %. But the probability of entry strongly increases if this topic is related to the local scientific structure. In fact, the coefficient for relatedness is equal to 0.00299. This estimate implies that an increase of 10% in the degree of relatedness is associated with a 66% increase ($[10*0.00299]/0.04534=0.659$) in the probability of entry. In a second model (model 2) we estimate the impact of the scientific importance of cities (number of publications in which the city is mentioned, lagged by one period) and topics (number of publications in which the title word appears, lagged by one period) on the emergence of new scientific topics in cities. Both variables have a positive and significant impact on the probability of entry, so we included these variables together with relatedness in a third model. Model 3 presents the results for the estimation of the full Eq. (3). All three variables positively and significantly influence scientific knowledge dynamics at the city level. With control variables included at the city and topic level, the coefficient for relatedness is slightly lower than in model 1 (0.00224) but it still has a substantial impact. In this specification, an increase of 10% in the degree of relatedness is associated with a 49% increase ($[10*0.00224]/0.04534=0.494$) in entry probability.

Table 5. The emergence of new scientific topics in cities (1989-2008)

Dependent variable is: Entry_t	Model 1 [Pooled OLS]	Model 2 [Pooled OLS]	Model 3 [Pooled OLS]
Intercept	0.04534*** (0.00165)	0.04534*** (0.00189)	0.04534*** (0.00159)

⁸ The econometric results, however, are robust to generalized linear models specifications (logit and probit regressions).

⁹ To compute the adjusted standard errors, we use the two-way clustering function written for the R statistical software by Arai (2011).

Relatedness _{t-1} (%)	0.00299*** (0.00016)	-	0.00224*** (0.00016)
Log (publications) _{t-1} [topic level]	-	0.01099*** (0.00076)	0.00996*** (0.00074)
Log (publications) _{t-1} [city level]	-	0.00524*** (0.00055)	0.00173*** (0.00039)
No. of observations	1,089,104	1,089,104	1,089,104

Notes: The dependent variable *entry* = 1 if a given topic (n = 1028) enters in the scientific portfolio of a given city (n = 276) during the corresponding 5-years window (n = 4), and 0 otherwise. Coefficients are statistically significant at the *p < 0.10; **p < 0.05; and ***p < 0.01 level. Heteroskedasticity-robust standard errors (clustered at the city and topic level) in parentheses.

The econometric results presented in table 5 indicate that relatedness plays a key role in the emergence of new scientific topics in cities over time. But to have a fully fledged understanding of knowledge dynamics, we also have to account for the process of creative destruction in the evolution of scientific fields. Therefore we now turn into the analysis of the exit of topics in cities by estimating Eq. (4), where $EXIT_{i,c,t} = 1$ if a topic i exits from the set of scientific topics of city c in time t , and 0 otherwise;

$$P(EXIT_{i,c,t}) = \beta_0 + \beta_1(RELATEDNES S_{i,c,t-1}) + \beta_2(CITY_{c,t-1}) + \beta_3(TOPIC_{i,t-1}) + \varepsilon_{i,c,t} \quad (4)$$

Table 6 presents the results for the estimation of Eq. (4). The probability of exit when all the variables are set to their means is given by the intercept $\beta_0 = 0.68501$. A probability of exit of 68% reflects the high instability of scientific topics, which is in part induced by the fine-grained classification and the large sample of cities and topics we analyze. But it is also true that scientific activities are more volatile than economic activities, as they can more easily migrate from one city to another with the mobility of academic researchers for instance. The exit of scientific topics is, however, strongly constrained by their degree of relatedness with the scientific portfolio of cities. As it can be seen from model 1 and model 3, relatedness has a negative and significant impact on the probability of exit of scientific topics. In the simplest specification (model 1), the probability of exit of a given topic in a given city during the next period decreases by about 5 % ($[10*(-0.00333)]/0.68501 = -0.048$) if the level of relatedness increases by 10 percent. Interestingly, the impact of relatedness is even higher when controls at the city and topic level are included. When variations in terms of scientific importance of

cities and topics are accounted for (model 3), an increase of 10% in the degree of relatedness is associated with a 10% decrease ($[10*(-0.00659)]/0.68501=-0.096$) in the probability of exit.

Table 6. The exit of existing scientific topics in cities (1989-2008)

Dependent variable is: Exit_t	Model 1 [Pooled OLS]	Model 2 [Pooled OLS]	Model 3 [Pooled OLS]
Intercept	0.68501*** (0.01391)	0.68501*** (0.00717)	0.68501*** (0.00563)
Relatedness _{t-1} (%)	-0.00333*** (0.00041)	-	-0.00659*** (0.00026)
Log (publications) _{t-1} [topic level]	-	-0.12034*** (0.00614)	-0.14998*** (0.00429)
Log (publications) _{t-1} [city level]	-	-0.08224*** (0.00520)	-0.02901*** (0.00473)
No. of observations	45,808	45,808	45,808

Notes: The dependent variable *exit* = 1 if a given topic (n = 1028) exits from the scientific portfolio of a given city (n = 276) during the corresponding 5-years window (n = 4), and 0 otherwise. Coefficients are statistically significant at the *p < 0.10; **p < 0.05; and ***p < 0.01 level. Heteroskedasticity-robust standard errors (clustered at the city and topic level) in parentheses.

The econometric results show that relatedness is a key driver of knowledge dynamics in science at the city level. We observed a strong and robust pattern of path and place dependence where scientific topics systematically tend to emerge in cities where scientifically related topics already exist, while relatedness in cities also tend to prevent the exit of topics. Since the city-topic relatedness variable combines relatedness between topics with the scientific expertise of cities, it is important to highlight that the cities that are specialized in the most central topics¹⁰ are more likely to make new discoveries and thrive scientifically speaking. In the next section, we discuss some implications of these findings.

5. Conclusion and discussion

This paper investigated the impact of scientific relatedness on knowledge dynamics in biotech at the city level worldwide during the period 1989-2008. We assessed the extent to which the emergence of new research topics and the disappearance of existing topics in cities are

¹⁰ A central topic is a topic that is scientifically related to many other topics. Degree centrality of topics can be computed from the topic-topic relatedness matrices described in section 3.

dependent on their degree of relatedness with existing topics in cities. Inspired by the product space framework developed by Hidalgo et al. (2007), we mapped for this purpose a scientific space, which is a network-based representation of scientific relatedness between topics. The links that connect topics are constructed from their co-occurrences within the titles of journal articles, and they express similarities in the capability requirements. As expected, we found strong evidence that new scientific topics in biotech tend to emerge systematically in cities where scientifically related topics already exist, while existing topics were more likely to disappear from a city when these were weakly related to the scientific portfolio of the city.

Of course, these findings raise many new issues that need more attention in future research. For instance, one wonders what the potential impact of extra-regional linkages could be on the rise and fall of scientific topics, as biotech researchers also collaborate and co-publish with other researchers outside their city to an increasing extent. Co-author and research collaboration networks are considered key channels of knowledge diffusion and learning among researchers. Moreover, the extent to which city-city networks matter for developing new research topics, may also depend on the degree of relatedness among the network partners involved (Phene and Tallman, 2002). Studies on networks suggest that more radically new knowledge is developed when actors bring in related competences (Nooteboom, 2000; Gilsing et al., 2007). Thus, non-local networks may provide relational access to the adjacent set of possible topics cities may branch into, next to the global and local context that give access to the adjacent possible. This needs to be taken up in future research.

Another key topic is how to measure scientific knowledge dynamics. In this paper, we used title words in scientific biotech publications to identify major cognitive developments within the field of biotech, and we analyzed knowledge dynamics at the city level by looking at the entry and exit of topics in the scientific topic portfolio of 276 biotech cities in the world. However, the use of title words is not unproblematic (Heimeriks and Boschma, 2013). Therefore, it remains a big challenge whether our findings will be affected by the use of alternative indicators, like the identification of major new cognitive events in the history of biotech science by leading experts or by means of the study of the evolution of patent activity in various technology classes. Patent analysis would also enable to construct a relatedness indicator by means of co-occurrence analysis based on the frequency of technology classes in the same patent document (Boschma et al., 2012).

Moreover, our main finding that new scientific topics emerge systematically in cities with related topics suggests that these are all recombinations of existing knowledge. While this

may be true for a large fraction of new topics, it is not necessarily true for all scientific topics. In fact, it is highly plausible that some new topics were true novelties, i.e. they had few or no related topics at the time they emerged. Krafft et al. (2011) used patent documents to identify a number of discontinuities in the evolution of the knowledge base in biotech, such as the rise of a new generation of biotech linked to bio-informatics. Making use of social network tools, they captured both genuine novelty through the emergence of completely new nodes (as proxied by new patent classes) and recombinations through the creation of new links between existing nodes. From a geographical perspective, it would be interesting to test whether novelties are more likely to enter in new or highly diversified biotech cities, and whether recombinations occur more often in more specialized biotech cities.

Furthermore, as the scientific space shows explicitly which topics require similar capabilities to the ones a city already has, it may help to inform researchers and policymakers about promising areas of further scientific development, comparative advantages and missing competences. This is an especially pressing issue in emerging sciences such as biotechnology, that are characterized by rapid growth, divergent dynamics, and new complementarities creating the need for wide-ranging cross-disciplinary competences (Bonaccorsi, 2008; Heimeriks and Leydesdorff, 2012).

Our results show that research strategies need to take into account existing local skills and infrastructures to create future capability and comparative advantage. This implies that policies that aim to develop biotech in cities from scratch are bound to fail. If policy wants to push new scientific fields in cities, it is better to connect to locally available scientific fields from which they can draw related capabilities. Topics (and fields in general) also differ in the number and specific nature of the capabilities they require, as cities differ in the number and nature of capabilities they have. Topics that require more capabilities will be accessible to fewer locations (as is the case in most topics in biotech), while locations with a wider range of capabilities will contribute to more topics. Biotechnology is characterized by a high level of branching into related knowledge, and as a consequence, policy should focus on developing a narrow set of related research activities in order to yield greater innovative output.

6. References

Amin, A., Cohendet, P., 2000. Organisational learning and governance through embedded practices. *Journal of management and governance* 4 (1), 93-116.

- Antonelli, C., 1995. *The economics of localized technological change and industrial dynamics*. Dordrecht: Kluwer Academic Publishers.
- Arai, M., 2011. Cluster-robust standard errors using R. Department of Economics and SULCIS, Stockholm University.
- Asheim, B.T., Coenen, L., 2005. Knowledge bases and regional innovation systems. Comparing Nordic clusters. *Research Policy* 34 (8), 1173-1190.
- Asheim, B.T., Gertler, M., 2005. The Geography of Innovation: Regional Innovation Systems, in Fagerberg, J., Mowery, D., and Nelson, R. (eds.), *The Oxford Handbook of Innovation*. Oxford University Press, Oxford, 2005, pp. 291-317.
- Asheim, B.T., Isaksen, A., 2002. Regional innovation systems. The integration of local 'sticky' and global 'ubiquitous' knowledge. *The Journal of Technology Transfer* 27 (1), 77-86.
- Atkinson, A.B., Stiglitz, J.E., 1969. A new view of technological change. *The Economic Journal*, 573-578.
- 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 (1), 31-56.
- Bonaccorsi, A., 2008. Search Regimes and the Industrial Dynamics of Science. *Minerva* 46, 285-315.
- Boschma, R., Balland, P.A., Kogler, D., 2012 Relatedness and Technological Change: The Emergence of New Technologies in U.S. Cities from 1981 to 2010. Paper presented at the Regional Studies Association European Conference, May 13-16, Delft, The Netherlands.
- Boschma, R., Frenken, K., 2011. Technological relatedness and regional branching, in: Bathelt, H., Feldman, M.P. and Kogler, D.F. (eds.) *Beyond Territory. Dynamic Geographies of Knowledge Creation, Diffusion and Innovation*. Routledge, London and New York, pp. 64-81.
- Boschma, R., Minondo, A., Navarro, M., 2013. The emergence of new industries at the regional level in Spain. A proximity approach based on product-relatedness, *Economic Geography*, 89 (1), 29-51.
- Boschma, R.A., 2004. Competitiveness of regions from an evolutionary perspective. *Regional Studies* 38 (9), 1001-1014.
- Breschi, S., Lissoni, F., 2009. Mobility of skilled workers and co-invention networks: an anatomy of localized knowledge flows. *Journal of Economic Geography* 9 (4), 439-468.
- Capello, R., 1999. Spatial transfer of knowledge in high technology milieux: Learning versus Collective Learning Processes. *Regional Studies* 33 (4), 353-365.
- Carvalho, R., Batty, M., 2006. The geography of scientific productivity: scaling in US computer science. *Journal of Statistical Mechanics* 10, 1-11.

- Cohen, W.M., Levinthal, D.A., 1989. Innovation and Learning: The two faces of R&D. *The Economic Journal* 99, 569-596.
- Cooke, P., 2006. Global bioregions: Knowledge domains, capabilities and innovation system networks. *Industry and Innovation* 13, 437-458.
- Cronin, B., Snyder, H.W., Rosenbaum, H., Martinson, A., Callahan, E., 1998. Invoked on the Web. *Journal of the American Society for Information Science* 49 (14), 1319-1328.
- Dosi, G., 1982. Technological paradigms and technological trajectories. A suggested interpretation of the determinants and directions of technical change. *Research Policy* 29 (4-5), 497-529.
- Eriksson, R.H., 2011. Localized spillovers and knowledge flows – How does proximity influence the performance of plants. *Economic Geography*, 127-152.
- Foray, D., David, P.A., Hall, B., 2009. Smart Specialisation: The concept. *Knowledge Economists Policy Brief* 9.
- Frenken, K., Hardeman, S., Hoekman, J., 2009. Spatial scientometrics. Towards a cumulative research program. *Journal of Informetrics* 3, 222-232.
- Gertler, M.S., 2003. Tacit knowledge and the economic geography of context, or the undefinable tacitness of being (there). *Journal of Economic Geography* 3, 75-99.
- Gertler, M.S., Levitte, Y.M., 2005. Local nodes in global networks: The geography of knowledge flows in biotechnology innovation. *Industry and Innovation* 12, 487-507.
- Gilsing, V.B., Nooteboom, B., Vanhaverbeke, W., Duysters, G., van den Oord, A., 2007. Network embeddedness and the exploration of novel technologies. Technological distance, betweenness centrality and density. *Research Policy* 37, 1717-1731.
- Gittelman, M., 2007. Does geography matter for science-based firms? Epistemic communities and the geography of research and patenting in biotechnology. *Organization Science* 18 (4), 724-741.
- Hardeman, S., 2012. The distributed organization of science, with empirical applications from the field of diabetes medicine. Eindhoven University of Technology, Eindhoven.
- Hausmann, R., Klinger, B., 2007. The Structure of the Product Space and the Evolution of Comparative Advantage. CID Working Paper 146, Center for International Development, Harvard University.
- Heimeriks, G., 2012. Interdisciplinarity in biotechnology, genomics and nanotechnology. *Science and Public Policy*, 1-16. doi:10.1093/scipol/scs070
- Heimeriks, G., Boschma, R., 2013. The path- and place-dependent nature of scientific knowledge production in biotech. *Journal of Economic Geography*, doi:10.1093/jeg/lbs052
- Heimeriks, G., Leydesdorff, L., 2012. Emerging Search Regimes: Measuring Co-evolutions among Research, Science, and Society. *Technology Analysis and Strategic Management* 24(1), 51-67.

- Heiner, R.A., 1983. The origin of predictable behaviour. *The American Economic Review* 73, 560-595.
- Hidalgo, C.A., Klinger, B., Barabasi, A.L., Hausmann, R., 2007. The product space conditions the development of nations, *Science* 317 (5837), 482-487.
- Hoekman, J., 2012. Science in the age of globalization, The geography of research collaboration and its effect on scientific publishing. Eindhoven University of Technology, Eindhoven.
- Hoekman, J., Frenken, K., Tijssen, R., 2010. Research collaboration at a distance: changing spatial patterns of scientific collaboration within Europe. *Research Policy* 39, 662-673.
- Hoekman, J., Frenken, K., van Oort, F., 2009. The geography of collaborative knowledge production in Europe. *Annals of Regional Science* 43 (3), 721-738.
- Katz, J.S., 1994. Geographical proximity and scientific collaboration. *Scientometrics* 31 (1), 31-43.
- Kauffman, S.A., 1995. *At Home in the Universe. The Search for the Laws of Self-Organization and Complexity*. New York: Oxford University Press.
- Klepper, S., 2007. Disagreements, spinoffs, and the evolution of Detroit as the capital of the U.S. automobile industry. *Management Science* 53 (4), 616 - 631.
- Krafft, J., Quatraro, F. Saviotti, P., 2011. The knowledge-base evolution in biotechnology: a social network analysis. *Economics of Innovation and New Technology* 20 (5), 445-475.
- Laudel, G., 2005. Migration currents among the scientific elite. *Minerva* 43, 377-395.
- Leydesdorff, L., 1989. Words and Co-Words as Indicators of Intellectual Organization. *Research Policy* (18), 209-223.
- Leydesdorff, L., 2008. On the normalization and visualization of author Cocitation data: Salton's cosine versus the Jaccard index. *Journal of the American Society for Information Science and Technology* 59(1), 77-85.
- Leydesdorff, L., 2010. The Knowledge-Based Economy and the Triple Helix Model. *Annual Review of Information Science and Technology* 44, 367-417.
- Leydesdorff, L., Cozzens, S., 1993. The delineation of specialties in terms of journals using the dynamic journal set of the Science Citation Index. *Scientometrics* (26), 133-154.
- Leydesdorff, L., Heimeriks, G., 2001. The Self-Organization of the European Information Society: The case of "Biotechnology". *Journal of the American Society for Information Science and Technology* 52 (14), 1262-1274.
- Leydesdorff, L., Persson, O., 2010. Mapping the Geography of Science: Distribution Patterns and Networks of Relations among Cities and Institutes. *Journal of the American Society for Information Science and Technology* 61 (8), 1622-1634.

- Martin, R. and Moodysson, J., 2012. Comparing knowledge bases. On the geography and organization of knowledge sourcing in the regional innovation system of Scania, Sweden, *European Urban and Regional Studies*, forthcoming.
- Martin, R., 2012. Knowledge bases and the geography of innovation. Lund: CIRCLE, Lund University.
- Maskell, P., Malmberg, A., 1999. The competitiveness of firms and regions. Ubiquitification and the importance of localized learning. *European Urban and Regional Studies* 6 (1), 9-25.
- Moodysson, J., 2007. Sites and modes of knowledge creation. On the spatial organization of biotechnology innovation. Lund: CIRCLE, Lund University.
- Neffke, F., 2009. Productive Places. The Influence of Technological Change and Relatedness on Agglomeration Externalities. PhD thesis, Utrecht University, Utrecht.
- Neffke, F., Henning, M., Boschma, R., 2011. How do regions diversify over time? Industry relatedness and the development of new growth paths in regions. *Economic Geography* 87, 237-265.
- Nelson, R.R., Winter, S.G., 1982. *An Evolutionary Theory of Economic Change*. Cambridge (MA) and London: The Belknap Press.
- Nooteboom, B., 2000. *Learning and innovation in organizations and economies*. Oxford, Oxford University Press.
- Phene, A., Tallman, S., 2002. Knowledge flows and geography in biotechnology, *Journal of Medical Marketing. Device, Diagnostic and Pharmaceutical Marketing* 2 (3), 241-254.
- Storper, M., 1995. The resurgence of regional economies, ten years later: The region as a nexus of untraded interdependencies. *European Urban and Regional Studies* 2 (3), 191 - 221.
- Strambach, S. and B. Klement, 2012. Cumulative and combinatorial micro-dynamics of knowledge. The role of space and place in knowledge creation. *European Planning Studies*, forthcoming.
- Todtling, F., Trippel, M., 2005. One size fits all?: Towards a differentiated regional innovation policy approach. *Research policy* 34 (8), 1203-1219.
- Trippel, M., 2009. Islands of innovation and internationally networked labor markets: magnetic centers for star scientists?, SRE-Discussion Paper2009/06.
- UNESCO, 2010. *UNESCO Science Report 2010. The Current Status of Science around the World*. UNESCO publishing: Paris.
- Van Den Besselaar, P., Heimeriks, G., 2006. Mapping research topics using word-reference co-occurrences: A method and an exploratory case study. *Scientometrics* 68 (3), 377-393
- Van Eck, N.J., Waltman, L. 2009. How to normalize cooccurrence data? An analysis of some well-known similarity measures. *Journal of the American Society for Information Science and Technology* 60(8), 1635-1651.

Wagner, C., 2008. *The New Invisible College: Science for Development*. Brookings Institution Press.

Wooldridge, J.M., 2003. Cluster-sample methods in applied econometrics. *American Economic Review* 93, 133-138.

Wuchty, S., Jones, B.F., Uzzi, B., 2007. The increasing dominance of teams in production of knowledge. *Science* 316 (5827), 1036-1039.

Zucker, L., Darby, M., 2007. Star scientists, innovation and regional and national immigration. National Bureau of Economic Research Working Paper, no. 13547, october 2007.