Journal of Informetrics 8 (2014) 259–272

Contents lists available at ScienceDirect

Journal of Informetrics

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

Analysing knowledge capture mechanisms: Methods and a stylised bioventure case

Thomas Gurney[a,*], Edwin Horlings[a], Peter van den Besselaar[b], Koichi Sumikura[c], Antoine Schoen[d], Patricia Laurens[d], Daniel Pardo[e]

[a] Rathenau Institute, Science System Assessment, Anna van Saksenlaan 51, 2593 HW The Hague, The Netherlands
[b] VU University Amsterdam, Network Institute & Department of Organization Science, De Boelelaan 1105, Amsterdam, The Netherlands
[c] GRIPS – National Graduate Institute for Policy Studies, 7-22-1 Roppongi, Minato-ku, Tokyo 106-867, Japan
[d] Université Paris-Est, ESIEE – LATTS – IFRIS, 2, bd Blaise Pascal, Noisy le Grand 93160, France
[e] CNRS – Aix-Marseille Université, LEST UMR 7317, 35 Avenue Jules Ferry, 13626 Aix en Provence Cedex 01, France

ARTICLE INFO

Article history:
Received 20 September 2013
Received in revised form 9 December 2013
Accepted 17 December 2013

Keywords:
Absorptive capacity
Knowledge transfer
Concept clusters
Non-patent literature references
Patent applications

ABSTRACT

Knowledge transfer between science and technology has been studied at micro- and macro-levels of analysis. This has contributed to the understanding of the mechanisms and drivers, but actual transfer mechanism and process, be they through codified or tacit sources, have very rarely been mapped and measured to completeness and remain, to a large extent, a black box. We develop a novel method for mapping science–technology flows and introduce 'concept clusters' as an instrument to do so. Using patent and publication data, we quantitatively and visually demonstrate the flows of knowledge between academia and industry. We examine the roles of exogenous and endogenous knowledge sources, and of co-inventors and co-authors in the application of university-generated knowledge. When applied to a stylised case, we show that the method is able to trace the linkages between base knowledge and skill sets and their application to a technology, which in some instances span over twenty-five years.

1. Introduction

Knowledge transfer between universities and firms has become increasingly institutionalised (Geuna & Muscio, 2009) as universities look for novel, more insightful, ways to enhance their economic and societal value through new technology spin-offs or start-ups (Audretsch, Lehmann, & Warning, 2005; Tijssen, 2006). Much of the previous literature has focussed on the facilitating actions and conditions for knowledge transfer such as scientific publications, conferences, informal interactions, collaborative and contract research, IP licensing, personnel exchanges and hiring – each with varying significance for industry (Ponomariov & Boardman, 2012).

A major challenge to evaluating these knowledge transfer routes and mechanisms is that of uncovering meaningful linkages between the technological outputs and scientific inputs. Knowledge transfer occurs most often at both the codified and tacit level, and the transfer processes and motivations within academic research versus those in industry settings are complex and evolving. However, what is not discussed in detail in the extant literature is the demarcation and measurement

* Corresponding author. Tel.: +31 62 4494529.
E-mail addresses: thomasgurney@outlook.com, thomasgurney@outlook.com, t.gurney@rathenau.nl (T. Gurney), e.horlings@rathenau.nl (E. Horlings), p.a.a.vanden.besselaar@vu.nl (P. van den Besselaar), sumikura@grips.ac.jp (K. Sumikura), a.schoen@esiee.fr (A. Schoen), laurensp@esiee.fr (P. Laurens), daniel.pardo@wanadoo.fr (D. Pardo).

1751-1577/$ – see front matter
http://dx.doi.org/10.1016/j.joi.2013.12.007

of the knowledge that is transferred (Bozeman, 2000). This is of utmost importance as the facilitation of transfer has been investigated but the question of *whether* knowledge has been transferred can only be answered by (a) being able to demarcate the object of transfer, and (b) measuring its point of inception, evolutionary path and eventual application.

Specific quanta of knowledge evolve along developmental paths, shaped by not only the scientific and technological developments of the laboratory in which it was conceived, but also by the further learning and skill sets of the scientists and inventors involved. By exploring the routes of codified knowledge transfer from inception to exploitation, we can begin to understand the processes and mechanisms of knowledge transfer. These include interactive knowledge production, the role a scientist's skill set plays, the effect of a scientist's peers – be they in the university or in the lab – and the transformative nature of science itself.

Whilst there have been substantive efforts at examining the facilitation processes and end-utilisation in an isolated sense – in that each step in the overall development process of a technology is analysed – a novel methodological approach is necessary to address the question of whether the whole transfer process has occurred. As mentioned previously this requires a method to both demarcate and track specific quanta of knowledge. In doing so, a clear view on the effectiveness of the facilitating conditions is possible.

In this paper we use and modify available tools and models, and integrate those with newly developed tools to provide a clearer and detailed picture of knowledge transfer, from start to finish. This paper starts with a discussion of the role of the scientist/entrepreneur, and that of his surroundings, in developing the necessary skill sets and knowledge for eventual transfer and application in industry. We then translate these insights into our methodology, which is described in detail. A crucial step in the methodology is the introduction of the idea of 'concept clusters', which refers to a small, cognitively cohesive agglomeration of scientific peer-reviewed publications. As an illustration, we apply the methodology briefly on a case. In the conclusions, we summarise the potential benefits, open methodological issues, and routes for further research.

2. Conceptual framework

The codification of knowledge takes two primary forms, patents and scientific publications. The use of patents as indicators was pioneered by Schmookler (1966) with many applications following (for examples, Schmoch (1993) and Fleming (2001)). However, many aspects of their indicator-oriented uses do have drawbacks (Pavitt, 1998), for example, not all innovations are patented (Arundel, 2001; Arundel & Kabla, 1998) with many innovations kept under a veil of secrecy (Brouwer & Kleinknecht, 1999) leading to underestimation of innovative potential or capacity. Analyses using patent indicators are typically based on the meta-data found in patents. Title words, abstract words and keywords (Courtial, Callon, & Sigogneau, 1993; Engelsman & van Raan, 1994), patent classifications (Leydesdorff, 2008; Tijssen & Van Raan, 1994), and patent/non-patent citations (Karki, 1997; Meyer, 2001) have all been used extensively. Many patent databases exist from which we extract the meta-data used in analyses, each with their own idiosyncratic advantages and disadvantages. These include disclosure requirements of prior art ('duty of disclosure'), wherein the USPTO requires an exhaustive list whereas the EPO requires a minimal listing. Differences also stem from the databases themselves, in terms of their formatting, whilst others relate to the practices of applying for patents through different national or supranational patenting offices. Despite the mentioned shortcomings, patents can be used for mapping knowledge transfer in a large part of the knowledge intensive economy because patent documents are highly detailed descriptions of the processes, applications and necessary information required for a technology. Citations within a patent document, either to other patent documents or scientific literature, add to this wealth of data. Patent documents encompass a wide range of technological fields and the major patenting offices (such as the USPTO or EPO) cover patent data from all countries (Tijssen, 2001).

Publications serve as the primary indicators for the defining characteristics and development of science. They are the most visible outcome of scientific endeavours, with an extensive range of indicators and methodologies developed. The analysis of publications shares a number of analytical approaches with patent analyses, such as word mapping (Callon, Courtial, & Laville, 1991) and citation analysis (Garfield & Welljams-Dorof, 1992; White & McCain, 1998). Using co-occurrences of combinations of words and cited references in publications is also becoming a common technique (Braam, Moed, & van Raan, 1991; van den Besselaar & Heimeriks, 2006).

The act of publishing itself is subject to a complex system of social and scientific norms, practices and reward systems (Merton, 1957). Publishing behaviours and patterns of scientists are governed in large part by these norms and practices, as well as by serendipity. The development of a university scientist's profile and portfolio are the result of search strategies (Horlings & Gurney, 2012) employed by the scientist. University-based scientists publish primarily to extend their professional and intellectual prowess, and regular publishing is considered a requirement. Industry based scientists are governed by similar constraints, and the firm benefits from publishing too – by becoming intimately involved with the basic science behind the technologies (Rosenberg, 1990), and their publications serve as a signal of their capabilities to the outside world (Hicks, 1995).

The conditions required for facilitating the development and transfer of knowledge depends heavily on the recipient knowledge platform. Knowledge assets (Nonaka, 1994), sector roles (Baba, Shichijo, & Sedita, 2009) and science-push and demand-pull concepts (Langrish, Gibbons, Evans, & Jevons, 1972), factor into a knowledge base's receptivity. In this manner – external knowledge sources, taking into account demand and current capabilities, are readily absorbed and entrained into stock knowledge bases and practices. This receptivity is known as 'absorptive capacity' (Cohen & Levinthal, 1990) and can best be described as "[t]he ability of a firm to recognize the value of new, external information, assimilate it, and apply it to commercial ends is critical to its innovative capabilities," (p.128). Key to this is the individuals involved, with the absorptive capacity of a firm tied to its constituent individuals' absorptive capacity, i.e. the right personnel are in place to take advantage of incoming information. As Cohen and Levinthal (1990) state, "Beyond diverse knowledge structures, the sort of knowledge that individuals should possess to enhance organizational absorptive capacity is also important. Critical knowledge does not simply include substantive, technical knowledge; it also includes awareness of where useful complementary expertise resides within and outside the organization" (p.133).

The use of patent and publication data, in the context of absorptive capacity, allows us to map knowledge inputs and outputs, and in doing so illuminate the mechanisms at work. The aim of this study is to provide a map of the cognitive route between the scientific origins and the technological output, with specific focus on the knowledge capture mechanisms operating. To this end, we have developed a method that shows:

1. *How the scientific background of the patent corpus links to the scientific output of the inventor* A patented product is the result of accretion over time of the research results, practices, skill sets and processes of the inventors involved. Visible in the patent documents are references to the underlying science (cited publications) and technology (cited patents) that were instrumental towards the development of the new (patented) technology. By linking the patent corpus and publication output, we can determine the background or necessary scientific requirements for the technologies.
2. *How the collaborative research environment of the researcher/inventor contributes to the development of the underlying science and to the developed technology*. Academic and industrial collaboration is common in high technology fields. Much of science is the result of collaborative efforts between researchers, where resources can be pooled and task allocation increases efficiency. As such, any contributions from a researcher's network will be visible in any publication authorship list or patent inventor list.
3. *What other knowledge is needed by the researcher/inventor for the development of a technology, and how this is appropriated.* A scientist must incorporate new results and skills from previous research done by others, to improve upon and modify their own intellectual prowess and breadth of skills. The absorptive capacity of the individual is measured by their entrance into, and adoption and integration of, new fields cited by the technologies they work in.

By mapping these three aspects of the knowledge stream, we clarify several of the mechanisms through which knowledge capturing is supported: (1) the own research by the researcher/inventor; (2) the collaboration network of the researcher/inventor; (3) the knowledge uptake process of the researcher/inventor.

To map aspect 1, we have developed an approach based on the overlap in content of the non-patent literature references (NPLR) found in the patent applications, and the publication corpus of the inventor. An individual publishes in multiple streams of research, with the streams being composed of publications highly similar to each other, which can be determined algorithmically. The similarity between the researcher/inventor's publications and the NPLR can be calculated so that the NPLR are co-located together with the research streams of the individual. Comparing the underlying total knowledge and skill set required to develop the technology with the knowledge and skills of the individual researcher/inventor shows the contribution of the latter to the technologies. The contributions of an individual's co-inventors and co-authors can be similarly constructed allowing us to map 2. Finally, to map aspect 3, a more refined approach is required. An individual's research streams may be broad in topic and time, and general statements can be made regarding the relevance and importance of an individual's knowledge and skills to a company's technologies. To examine the specific scientific fields the technology draws upon (as defined by the NPLR and what fields they originate from), we have developed a method focusing on the specific scientific concepts and methods necessary for the technologies described in the patent documents. By identifying the specific concepts utilised in the technologies and at which point in time the researcher/inventor develops or integrates them into their knowledge base, we are able to view from where, and from what original form, new knowledge assets are derived. This method utilises *concept clusters*, which are defined and operationalised in detail in the next section. Concept clusters are used to map aspect 3 and to examine the detailed concepts and methods in aspects 1 and 2.

2.1. Concept clusters

A broad description of an individual's knowledge and skill sets may be derived through examination of the titles used, references cited, keywords used (and more) in their publication corpus. By adding the NPLRs of the patent applications to the individual's publication corpus, we are able to discern which aspects of an individual's corpus are similar to the NPLR. To discern general research themes within the combined corpus, we utilise the Louvain clustering method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) on all nodes and edges with no similarity threshold i.e. all edges are considered by the algorithm. The algorithm optimises the modularity of a network, to identify macro-clusters in the network. The modularity of a network refers to the comparison of densities within a partition (or cluster/grouping) and densities between partitions. Modularity is calculated by comparing the actual distribution of nodes and edges within a network with a random distribution. If the modularity is positive, a non-random structuring within the network is expected. The method is based on moving nodes and their edges between different partitions and on re-calculating and comparing the modularity of the modified partitions. If there is a positive change in the modularity in the new partition, the node is retained in the newly formed partition. If change is negative, the node is assigned to its original partition. This is an iterative process, calculated for every node in the

Table 1
Concept cluster composition.

	Publication authored by	
	Inventor	Other
Cited by patent		
Yes	A	B
No	C	–

network (Girvan & Newman, 2002; Newman, 2004).[1] The presence of singletons in the corpus of publications is rare as the research areas of the inventor/author and of the NPLR are expected to be similar. Additionally, the research trajectories of the scientists included in this study conform to standard behaviours. Their research output is the result of informed strategies on the part of the scientists, resulting in output that forms a cohesive cognitive chain from early PhD research through to their current research (Gurney, Horlings, & Van Den Besselaar, 2012; Horlings & Gurney, 2012).

The meta-data occurrences in each cluster are then examined to identify the general themes of that macro-cluster. To identify *specific* topics, each macro-cluster is isolated and the same clustering algorithm is applied to produce micro-clusters. These micro-clusters constitute the immediate publication environment of the NPLRs. Depending on the variety of subjects in the publication corpus, macro-clusters can range in size from 10 to 100+ publications whereas each micro-cluster is typically no larger than 10 publications.

We refer to these micro-clusters or immediate environments as 'concept clusters'. The publications cited by the patent applications (NPLR) form the nucleus of the concept clusters of interest to this study and each concept cluster contains at least one NPLR. Surrounding this nucleus are the publications most similar in terms of title word and cited reference combinations (van den Besselaar & Heimeriks, 2006), and the borders of each concept cluster are algorithmically delineated into communities (Blondel et al., 2008). A concept cluster contains, in varying proportions, publications authored by the researcher/inventor (which can be cited or not by the patent applications), and publications written by others that are cited by the patent application. The specific composition of a concept cluster describes the knowledge utilised in the patent application, in terms of the knowledge base and skills internal or external to the researcher/inventor.

Table 1 illustrates the possible publication origin types – A, B and C – in a concept cluster. For each concept cluster, a mix of publication type can result. So a concept cluster may contain:

1. Type A publications – this indicates direct contributions by the inventor to the required concepts and skill sets. The research and concepts contained within the publication are either necessary for, or directly related to, the development of the technology.
2. Type B publications – we assume that some knowledge is outside the expertise of the inventor.
3. Type C publications – whilst the inventor is not cited directly, his publications are highly similar to the NPLR in the concept cluster hence the inventor possesses skill sets and background knowledge similar to the NPLR.

As defined previously, absorptive capacity is the ability to recognise, assimilate and integrate new knowledge, and apply it in a novel manner. The absorptive capacity of an individual who is an inventor of the technologies can be determined by analysing the similarity and presence (or lack thereof) of their contributions to the concept clusters. The greater the number of occurrences of their own publications that are similar to the NPLR, the higher the enabling potential for absorptive capacity (Cohen & Levinthal, 1990). Whilst this is not a guarantee of successful internalisation (Jansen, Van Den Bosch, & Volberda, 2005), it is a requirement.

Simultaneously, the timing of an inventor's publishing entrance into a concept cluster is important. If publications by an inventor appear (Type A or C) first, followed by NPLR (Type B), we conclude the inventor has previously developed the skill sets and knowledge required for the technologies. We argue this as follows: as required by most patenting offices, directly related literature or previous patent applications that the inventor(s) are involved in must be presented to the examiner.[2] I.e. if the inventor has published material directly related to the invention, it must be made known to the examiner.

Following this line of reasoning, if NPLR were published prior to the inventor's publications, we assume the inventor previously did not have the knowledge necessary to publish on the matter. By researching and preparing the technology described within the patent application, although the inventor may not have published previously on research necessary for the technology, they develop their knowledge in a 'learning by doing' manner. This is common in many industrial sectors (Arora, 1997; Lieberman, 1987). It is necessary for the successful implementation of R&D (Sagar & Van der Zwaan, 2006),

[1] For a more detailed explanation of clustering algorithms in general, see Palla, Derényi, Farkas, and Vicsek (2005) For a comparative analysis of Blondel et al's algorithm versus others' see Lancichinetti and Fortunato (Lancichinetti & Fortunato, 2009).

[2] USPTO Manual of Patent Examining Procedure Section 704.10 Requirements for Information; and Part B, Chapter 3, Section 2. Scope of Search in the EPO Guidelines for Examination.

and a key factor in absorptive capacity (Cohen & Levinthal, 1990). The outcome of this response to the knowledge gap is publications (and research) by the inventor that are highly similar to the original NPLR, and potentially relevant to future developments of the technology. How soon an inventor publishes after recognising the value of the knowledge referenced by the NPLR indicates the perceived importance of the knowledge to the firm as well as the individual, as well as his ability to address it. Additionally, through publishing (and the research involved in publishing) the firm also benefits by becoming intimately involved with the basic research (Rosenberg, 1990) and the publications act as a signal of the firm's capabilities to the outside world (Hicks, 1995). Through this reasoning, we conclude that the increasing similarity between the NPLR and IA's publications indicates the increase of absorptive capacity by IA in the subject matter.

Using this approach, we can map aspect 3 from the previous section, and add to the first two. We determine if an inventor is a leader in the production of knowledge necessary for the technologies, or a follower. If a follower, does the inventor incorporate the necessary knowledge and skill sets into their portfolio early, demonstrating a high level of absorptive capacity? If necessary knowledge and skills lie outside the portfolio of the inventor, do his collaborators provide any of the knowledge or skills?

3. Previous work

Previous studies typically utilise text-mining approaches or citation matching to provide a linkage between patents and publications. Text-mining approaches generally involve methodologies that identify topical clusters in patents and publications using words (title, abstract, or full text) and link the two corpora together through the similarities between the topical clusters. Mogoutov, Cambrosio, Keating, and Mustar (2008) use a combinatorial approach to map innovation in the biomedical field of microarrays. Relevant concepts are extracted from multiple datasets, namely those of publications, patents, and research project data. A matching algorithm links the datasets through their shared concepts. They specifically try to avoid using pre-determined topic areas or research areas, to allow some qualitative room for interpretation after the matching has been completed. They successfully demonstrate a link between, and within, scientific fields through shared concepts.

Magerman, Van Looy, and Song (2010) provide a very thorough review of the state of the art of the text-mining approach. In addition, their study tests the effectiveness of distance measures when linking patents and publications via text-mining. With only 30 patents and 437 publications, Magerman et al. use a smaller dataset than what is typically encountered (these are notable figures as oft-used similarity distance measures rely on large datasets to provide higher-quality matching outcomes.) Magerman et al. acknowledge this and conclude that the overall number of records would likely increase the chance of linking patents and publications.

Text mining can rely on an abundance of methods, which are highly variable and customisable. However, some limitations of text mining also become apparent. The different vocabularies employed between patents and publications pose a risk to accurate matching. The size of the sample may result in misleading or inaccurate matching options. A further limitation is one of a changing vocabulary over time within a field of science. In publishing, the audience and indexer effects (Leydesdorff, 1989; Whittaker, 1989) may lead to fewer and fewer matches between publications and patents further apart in time. Text mining is typically a resource intensive approach, and requires extreme care due to the complex nature of linguistic behaviours and anomalies.

Citation matching is easier as it involves extracting the bibliographic non-patent literature references (B-NPLRs) from the patent documents and finding the corresponding twin in whichever publication database one uses. Unfortunately, the requirements for including citations (patent and non-patent) in patent applications vary drastically between patenting offices, with the duty of disclosure a prime example relevant to this study. The move to include in-text non-patent literature references (IT-NPLRs) is a recent development as the availability of extraction tools for full-text documents has increased. A study by Tamada, Naito, Kodama, Gemba, and Suzuki (2006) addresses the issue of IT-NPLRs, focussing specifically on Japanese patent documents. They argue that as there is no requirement by the Japanese Patent Office to include front-page references, patent output indicators that utilise only B-NPLR may miss relevant references to science. To counter this, they use references found in the text of the documents to successfully identify under-reported scientific fields cited by patent applications. They conclude that the inclusion of both in text and bibliographic citations enriched their data sets and provided balance between objective and strategic referencing of literature in patent applications.

Meyer (2002) examined the use of citations in patent and publication centred studies. He formed a typology of the most often used approaches including patent citation analyses, industrial scientific activities, and university and academic patenting. His critiques of the techniques essentially point to the misuse of analytical tools and methods from one field to another. He notes that techniques that use these approaches do not take into account basic characteristics of patenting and publishing. Some specific aspects include the propensity to publish varies across fields, citations can be negative or positive and publishing is not the only output of the laboratory. In terms of patenting, similar problem characteristics should be taken into account, such as patenting propensity across industries varies, not all inventions are patented, and a significant portion of patents are strategic, designed to block innovation in a competitor. Meyer may have examined these aspects over ten years ago, but the principles remain valid today when discussing methodologies using citation behaviours of publications and patents. Regarding NPLRs in patents, the relative abundance of references to scientific literature versus non-scientific literature is an indicator of the quality (Branstetter, 2005) and proximity to science (Callaert, Van Looy, Verbeek, Debackere, & Thijs, 2006) of the patent application. What is generally understood and accepted is that the presence of citations to scientific literature in patent documents indicates a cognitive link to, or awareness of, the related scientific concepts (Tijssen, 2001).

4. Method

The methodology we developed consists of various steps. The first step is to select the inventor/researchers that play a crucial role in the knowledge transfer case under study. The second step is the data-collection of the papers and patents of these individuals (Section 4.1). Then we do publication clustering (Section 4.2) and patent application clustering (Section 4.3). The next step is to link the patent applications and publication clusters, using a specific visualisation tool (Section 4.4). After having described the method, we demonstrate a proof of concept in Section 5.

4.1. Data collection

Many patent databases exist, each with their own idiosyncrasies, some of which stem from the databases themselves whilst others relate to the practices of different national or supranational patenting offices. In our study we use the PatSTAT database prepared and developed by the EPO, as it aggregates various other databases, and is considered one of the most extensive patent databases. For our publication data, we use the Thomson Reuter's Web of Science (WoS) as our primary source of publication data, supplemented by CV data from the scientists involved.

The sources and type of data come from:

1. Patent data – we extracted all patent applications with the selected inventors listed as an applicant from the EPO PatSTAT database along with all other inventor(s) data; this also included all patent applications with the firm under study listed as an inventor or assignee, and the selected inventors as assignee.
2. Publication data[3] – we extracted from WoS all publications with the inventors' firm listed as an institution; and all publications with the selected inventor listed as author.

These base data were parsed using SAINT[4] (Somers, Gurney, Horlings, and Van den Besselaar, 2009) and managed in a relational database. Further data were collected from the patents, specifically and where found:

1. Bibliographic NPLRs (B-NPLRs) – these are citations included primarily by the examiner and added as front-page references.
2. In-text NPLRs (IT-NPLRs) – citations to publications visible in the body of the patent. These IT-NPLRs were automatically extracted from the full-text versions of the patent documents by custom software.

All patent applications were then grouped according to their first filing, with the priority patent application representing the collective. Single-priority based families are collections of patent applications that claim a specific application as the first or priority application. The priority patent is included in the collection (Martinez, 2010). This is done to account for variations in NPLR reporting and inclusion across different patenting offices. A second reasoning is that any derivative applications are close extensions of the priority patent, thus one could expect the NPLRs from the collective to extend to the other applications in the group. Further references in this paper to these patent collectives use the term 'priority patent' to mean the *priority patent application representing the collective*.

The NPLRs were then normalised for search of their twin in WoS. If there were no NPLRs linked to any given priority patent, the NPLRs of derived citing patent applications (i.e. applications citing the original application as priority) were included. Both NPLR sets were parsed and, as far as possible, their WoS publication equivalents found. A manual check was performed to see if the retrieved documents matched the original NPLR. If any discrepancies in meta-data did not allow for a proper match, the affected records were not utilised in any further analysis. The verified documents were then parsed and processed together with the inventor's publications to create a master publication database and the origins of each document recorded.

4.2. Publication similarity and concept clusters

Publications are clustered by their shared combinations of title words and cited (van den Besselaar & Heimeriks, 2006). The degree of similarity is calculated using the Jaccard similarity coefficient. Clusters of publications were automatically assigned using the Louvain method of clustering (Blondel et al., 2008) grouping publications based on their degree centrality and relative weights of edges between nodes. These clusters are referenced further as 'research streams'. Each research

[3] English language only.
[4] SAINT (Science-system Assessment Integrated Network Toolkit – a Rathenau Instituut open-source software suite designed to parse, clean and organise bibliometric data to be later used in relational database software such as MS Access and MySQL.

Table 2
Summary of data collected for analysis.

Data	Feature	Count
Patent applications	Patent applications with FA listed as Assignee and IA as Inventor (2000–2008)	242 (115 Priority patent app.)
	INPADOC families	90
Publications	IA publications retrieved from WoS.	931 (786 pre-2009)

stream is then isolated and the community detection algorithm of Blondel et al. is run on the individual streams resulting in the concept clusters.

4.3. Patent clustering

Patent applications are grouped by INPADOC family ID and the NPLR of the INPADOC families within concept clusters are noted. The Jaccard similarity coefficient is calculated between INPADOC families using the shared concept clusters in which their NPLR occur, and the community detection algorithm of Blondel et al. is used to designate INPADOC clusters.

4.4. Visualising patents and publications

We have developed a method (Horlings & Gurney, 2012) that allows the specific research trails an individual has developed, to be visualised in a uniquely clear manner. We have built upon this method by adding patent applications, of which the individual researchers are listed as inventor, to their corpus of publications. The thematic and knowledge base aspects of the patents and publications are linked, not through direct citations by patents to the publications, but through shared thematic or topical research areas of the cited NPLR and the inventors' corpora of publications. Even if the patent document does not cite the individual's publication corpus directly, other cited literature may (or may not) cluster within the inventor's areas of expertise. This approach results in a tangible, visible, shared knowledge base between the patent and publication.

We arranged the patent applications and research streams along two axes, that of time on the x axis and research streams and patents on the y axis. Longitude is defined as [(year of publication) − (year of first documented publication in the dataset)/(range in years) × 360] − 180. Please note that the small clusters of papers visible within each research stream represent the annual production within that stream, and not the concept clusters – the latter contain papers published over a variety of years. Latitude is defined as [(stream number)/(total number of streams) × 180] − 90. The nodes were positioned with the GeoLayout in Gephi (Bastian, Heymann, and Jacomy (2009)), using an equirectangular projection.

5. Proof of concept

In this section we apply the methodology on a stylised case – stylised, as we are primarily interested in demonstrating that the methodology is able to map the knowledge streams, as well as mechanisms underlying these streams. The case is based on a real case, in which we aim to analyse in depth elsewhere (Gurney et al., 2013). The essential mechanics of the methodology are discussed here, with additional detail provided in the in-depth study. Images here are stylised, with data utilised to illustrate the necessary components.

5.1. Case selection

Our case study involves a prominent biotechnology researcher, who is strongly involved in cancer therapeutics at the firm he founded in 2001 and the university in which he is a professor. The researcher both publishes and patents extensively at his university and within the firm. The individual, firm and university shall further be referred to as, respectively, IA, F A,and UA. IA maintains direct links between his research at UA and research conducted at FA. This enables us to draw upon his extensive publishing history as well as his numerous patenting activities at both university and firm.

5.2. Data summary

Table 2 shows a summary of the various data collected. Patents cover the period 2000–2008[5], and the publications cover all publications in the categories defined in Section 4.1 up to 2011. The large number of patent applications (242) may not be typical of most companies in this field. The breadth of the patent applications, as exemplified by the number of INPADOC

[5] Patent applications up to 2008 were chosen as there is considered to be a delay in the completeness of patent data in PatSTAT. 2008 was chosen as the last year as we could be more certain that all possible patent data were included.

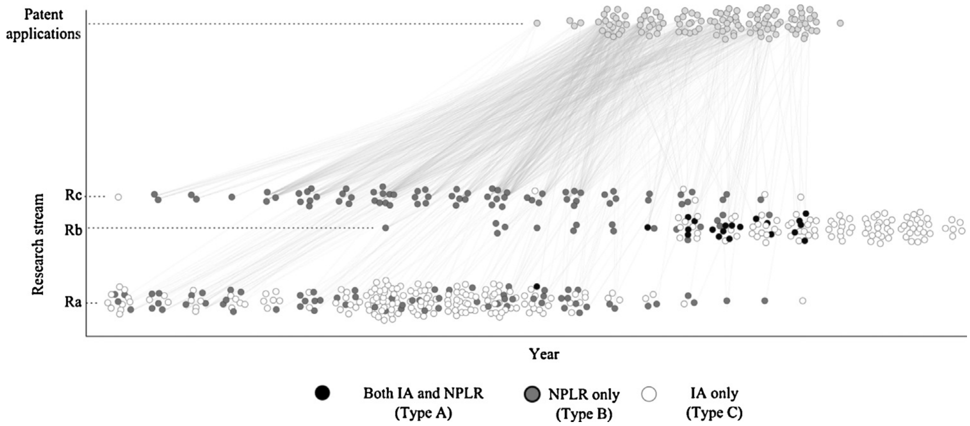

Fig. 1. Stylised image of patent applications and publications – authored by IA and/or cited by the patent applications over time. (Note: For type A, B, C, see Table 1).

families (90) is also large. IA is a prolific author with, in 2011, 931 publications to his name. This is an exceptionally high number and we assume many of these publications are purely the result of him being head of a large institute in which his name appears as author as a matter of seniority.

5.2.1. *Mapping the links of the scientific background of the patent corpus to the scientific output of the inventor*

Fig. 1 is a stylised image showing portions of the total corpus of IA publications and the NPLR of the patent applications differentiated into research streams. Noted in Fig. 1 are the publications authored by IA and publications cited by the patent applications over time. The visualisation was constructed as detailed in Section 4.4. For ease of reference, each research stream has been labelled **R*x*** (Ra, Rb, etc.) There are multiple research streams in our example, which refer to actual publications of our individual IA and the NPLR of the patent applications. Each stream has been selected to highlight different aspects of both the number of NPLR within each research stream, relative to publications of IA that are not NPLR as well as those authored by co-inventors of IA.

In Fig. 1, research stream **Rb** contains a considerable number of NPLR authored by IAas seen by the black nodes. NPLR not authored by IA form the beginning of the stream, and the stream becomes increasingly populated by publications of IA, both cited and not by the patent applications. Streams **Ra** and **Rc** contain NPLR not authored by IA, publications authored by IA but not cited and very few NPLR authored by IA, co-located in the same stream. Each stream may contain a mixture of publication types (A, B or C), and the proportional presence of IA's publications (cited or not) in the stream indicates the proximity of IA's research to the research cited by the patent applications.

In Table 3, the NPLR distribution in the patent applications is noted. Most of the NPLR come from within the text of the patent documents. 65 NPLR are found in both the text and bibliography of the applications. IA's publications make up 10% of the NPLR, with a proportionally larger number being cited in the bibliography.

From Fig. 1 and Table 3 we can conclude that aspects of research conducted by IA are relevant and necessary to the technologies represented by the patent applications. Some aspects are not within IA's expertise, such as those in research stream **Rc**. IA's research is cited in many instances in both the text of the document and the bibliography. Research that is cited but not authored by IA is often very similar to IA's publications, as seen in stream **Ra**.

Table 3
Summary of NPLR.

Data	Feature	Count
NPLRs	Unique NPLRs found (and matched to WoS) in all patent applications	525_a unique (2037 total)
	In B-NPLR only	147_b
	In IT-NPLR only	313_c
	In both B-NPLR and IT-NPLR	65_d
NPLR citations	Count of IA's publications cited in NPLR (Note: $\% = x/a$)	55 (10%)
	In B-NPLR only ($\% = x/b$)	18 (12%)
	In IT-NPLR only ($\% = x/b$)	19 (6%)
	In both B-NPLR and IT-NPLR ($\% = x/d$)	18 (27%)

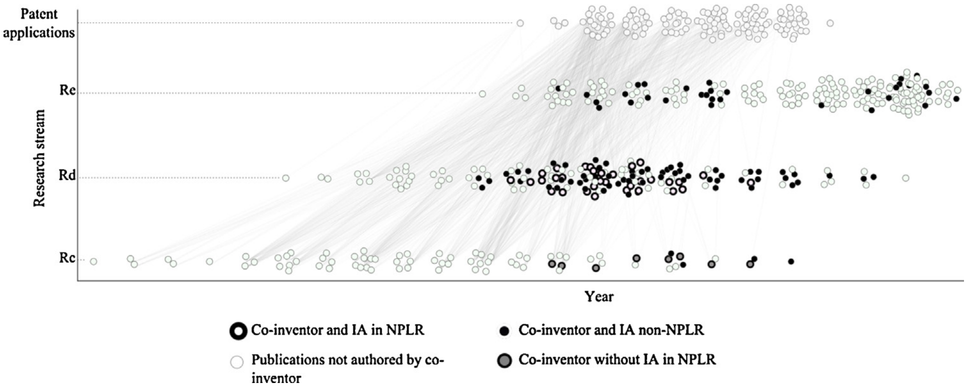

Fig. 2. Co-inventors of IA and their presence in NPLR and publications.

5.2.2. Mapping contributions of the inventor's research collaborations in the patent and publication data.

Fig. 2 is a stylised image of portions of the total corpus of IA publications and NPLR. The presence of co-inventors as authors of publications is noted. In research stream **Rc**, there are a number of NPLR authored by IA's co-inventors but do not feature IA as an author. In stream **Rd** there is a high proportion of NPLR and non-NPLR authored by both IA and his co-inventors. Stream **Re** contains many publications that are authored by both IA and his co-inventors but not cited by any of the patent applications. All three streams contain many papers by IA which were not co-authored by the co-inventors, and these papers are partly NPLR, and partly non-NPLR. The streams also contain NPLR not authored by IA or his co-inventors.

In Table 4, the presence of co-inventors publications as NPLR in the whole corpus is summarised. Only 9 publications written by IA's co-inventors (without IA as author) are cited by the patent applications in the NPLR (8 are shown in stream **Rc** in Fig. 2, the last is located in another stream) and most are IT-NPLR (in-text). The number of co-inventors' publications cited by patent applications (excluding publications co-authored by IA) is far lower than the number of IA's publications cited by the applications. IA's co-inventors appear as inventors without IA on 30 patent applications.

5.2.3. Mapping the inventor's level of adaptive knowledge use (absorptive capacity) necessary for the development of a technology

To demonstrate the absorptive capacity of IA in relation to the development of the technologies, the concept clusters and their utilisation by the patent applications are mapped over time. To construct the concept clusters, each research stream in the total corpus is isolated. The community detection algorithm of Blondel et al. (2008) is run over each isolated stream. Each resulting community or concept cluster contains a mixture of publications (Types A, B and/or C from Table 1), with at least one NPLR forming a nucleus. We grouped each patent application into INPADOC clusters, based on the similarities in of the aggregated main group IPC codes of the patent applications' parent INPADOC families. The resulting INPADOC clusters represent the different technologies in which IA is involved, and their development over time.

Figs. 3–5 show stylised representations of the appearance in time of the concept clusters containing NPLR cited by the INPADOC clusters. The concept clusters represent not only the immediate knowledge environment of the NPLR, but also the degree of similarity between the inherent skillsets of IA and the skillsets referenced by the patent applications in the NPLR. In this manner we can map the knowledge and skillsets necessary for the development of the technology (and therefore cited by the patents). IA's own work may not be cited in many instances but are highly similar to the NPLR and are co-located in the same concept cluster. This mapping strategy also shows at what stage of the technologies' development the skillsets and

Table 4
Summary of co-inventors' publication and patent application contributions.

Category	Feature	Count
Co-inventors		
Publications	Cited in NPLR (EXCLUDING publications co-authored by IA)	9
	In B-NPLR only (% = x/b)*	2 (1.5%)
	In IT-NPLR only (% = x/c)*	5 (1.5%)
	In both B-NPLR and IT-NPLR (% = x/d)*	2 (9%)
	Publications NOT cited by patent applications but co-authored with IA	251
Patent applications	FA patent applications without IA as inventor	30 (12%)

* *Note*: for b, c and d values see Table 2

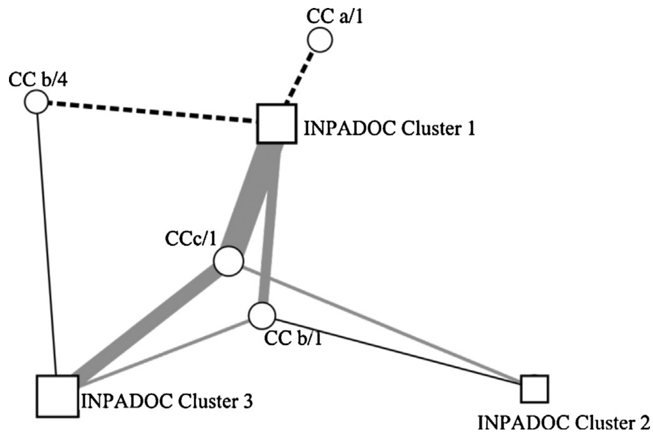

Fig. 3. Concept clusters cited by INPADOC clusters containing only NPLR not authored by IA (Type B publications). (Note: Concept clusters are labelled CC. Size of nodes indicates count of publications and count of INPADOC families. Thickness of edges indicates number of citing INPADOC families. Edge colour indicates at what phase in the age of the INPADOC cluster the concept is cited, grey = early, dashed = middle, black = late).

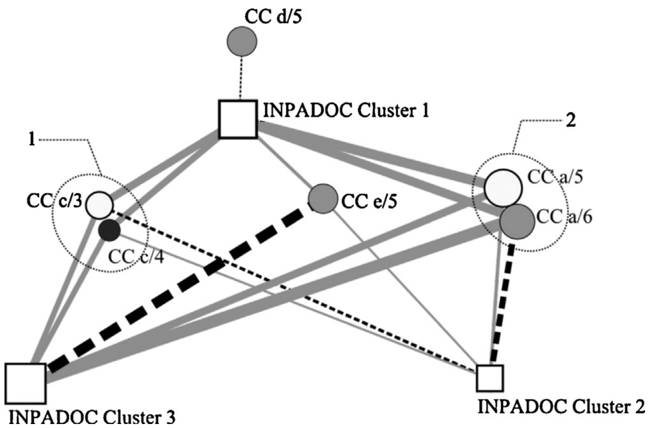

Fig. 4. Concept clusters cited by inpadoc clusters containing NPLR not authored by IA (Type B) and IA publications not cited by patent applications (Type C). (Note: Node size indicates count of publications and count of inpadoc families. Node shading indicates time of appearance of IA's publication into concept cluster, white = early, grey = middle, black = late. Edge thickness indicates number of citing INPADOC families in the INPADOC cluster. Edge colour indicates period of citation, grey = early, dashed = middle, black = late).

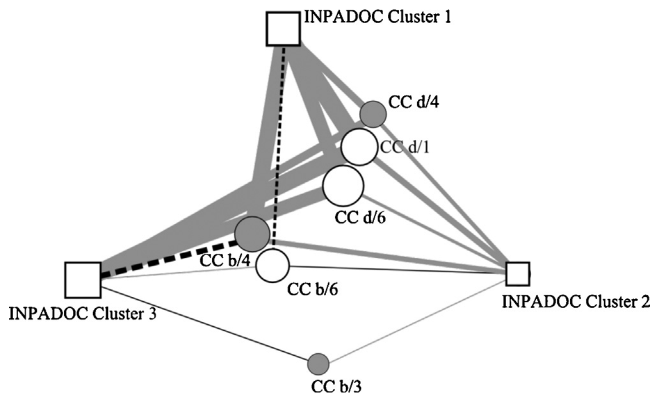

Fig. 5. Concept clusters cited by inpadoc clusters containing NPLR authored by IA (Type A), NPLR not authored by IA (Type B) and IA-authored publications not cited by patent applications (Type C). (Note: Concept clusters are labelled CC. Node size indicates count of publications and count of INPADOC families. Node shading indicates timing of IA's entrance into concept cluster, white = early, grey = middle, black = late. Edge thickness indicates number of citing INPADOC families in the INPADOC cluster. Edge colour indicates period of citation, grey = early, dashed = middle, black = late).

knowledge from outside IA's expertise are cited. The varied compositions of the concept clusters differ between Figs. 3–5, with detailed descriptions of the composition preceding each figure.

The nomenclature of the concept clusters refers to the original research streams in which the publications are located. In the examples presented in Figs. 3–5, concept clusters are represented by the term CC *x*/*y*. 'CC' is a general identifier for a concept cluster, with the subsequent *x*/*y* values referring to, for *x*, the larger stream in which the concept cluster is located, and *y* referring to the identifier of the sub-cluster, or sub-research stream in which the NPLR (and if present, the author's publications) are located. For example, CC *b*/4 come from research stream **Rb**, endocrinology related to porcine spinal cord and mouse-human models, and within that research stream – concept cluster 4 refers to the production of neuromedin-B and -U in the spinal cord of mammals.

The shading of edges between the INPADOC clusters and the concept clusters identify the period in the lifespan of the INPADOC clusters that patent applications start to cite publications in the concept clusters. Edge thickness indicates the total number of cites to publications in that concept cluster.

Fig. 3 includes concept clusters that contain only NPLR not authored by IA (Type B from Table 1). The skill sets and knowledge base embodied by these publications are considered to be outside the expertise and skill sets of IA as they do not contain any similar IA publications (in terms of title words or cited references).

In Fig. 3 the NPLRs located in the concept clusters are cited at different stages in the development of the technologies of INPADOC clusters 1–3. For many of the concept clusters cited in Figs. 3 and 4, the original research streams can be found in Fig. 2. For example CC *b*/4 (a concept cluster found in research stream **Rb** from Fig. 2) is cited in the middle development period by INPADOC cluster 1 (as signified by the dashed edge), but at a later period INPADOC cluster 3s development phase (as signified by the solid black edge) whereas CC *c*/1 (a concept cluster found in research stream **Rc** also from Fig. 2) is cited early (grey edge) in the development of all the INPADOC clusters. CC *a*/1 from **Ra**, is cited in the middle development phase by only INPADOC cluster 1.

Compared to Fig. 3 in which only areas of research not by IA are shown, Fig. 4 shows the concept clusters containing a mixture of NPLR not authored by IA (Type B) and publications authored by IA but not cited by the patent applications (Type C). Put another way, the concept clusters in Fig. 4 contain NPLR very similar to IA's work but do not cite him directly.

In the highlighted area 1 are two concept clusters (CC *c*/3 and *c*/4), both from research stream **Rc** in Fig. 2. The research contained within CC *c*/3 is cited in the early (grey edges) phases of development of INPADOC clusters 1, 2 and in the middle phase (dashed edge) of development by INPADOC cluster 2. IA's publications appear early in CC *c*/3 (as signified by the white node colour) but late in CC *c*/4 (as signified by the black node colour). This implies that the specific research within CC *c*/3 is necessary from an early stage in the development of the technologies in INPADOC clusters 1 and 2, and IA begins publishing in this topic at the same time it is cited. However, for CC *c*/4, although the research is cited early by INPADOC clusters 1 and 2, IA only begins publishing later in time on the topic. This implies IA recognises both the value of the research (in that he cites it in the patent applications) and the need to perform more research in the same topic, but also exhibits 'learning by doing' behaviour. He previously did not publish on the topic, but However, the same research is cited only in the middle development phase by INPADOC cluster and IA is publishing in this concept cluster also at an early stage, whereas IA only starts publishing in CC *c*/4 at a late stage in the concept cluster's lifespan. In highlighted area 2, concept clusters CC *a*/5 and *a*/6 from research stream **Ra** are cited in the early to middle development phase of the technologies. In CC *a*/5, IA's publications are found early in the lifespan of the concept cluster, but in CC *a*/6, his publications only appear in the middle stages of the concept cluster's lifespan.

Relating this to IA's absorptive capacity; IA has already published similar publications – taking into account the macro-clusters (considering CC *c*/3 and *c*/4 and CC *a*/5 and *a*/6 are from the same respective research streams) and highly similar publications considering each concept cluster separately. For the concept clusters in which IA's publications appear only in the middle or late phases, we conclude that IA has recognised the importance of the research being cited by the patent applications and, demonstrating a degree of absorptive capacity, begins to populate the concept clusters with his own publications (not cited by patents).

In Fig. 5, IA's direct knowledge and skill contributions to the technologies, as seen by the technologies citing IA authored publications in the concept clusters is more apparent. As also found in Fig. 4, the concept clusters are cited by the technologies at different stages of development. In Fig. 5 however, the concept clusters also contain NPLR authored by IA. In many cases, concept clusters containing NPLR authored by IA are cited during the early stages of the concept cluster lifespans, and others in the middle stages. Many of these come from the same research streams (for example CC *d*/4, *d*/1 and *d*/6 are from research stream **R*d***) and are cited by all three technologies. In the case of CC *d*/4, there are some transitive similarities to *d*/1 and *d*/6, and IA only begins publishing in the middle stages of that concept cluster's lifespan.

The absorptive capacity of IA is again demonstrated here as, in many cases, IA's research necessary to the technologies is cited early. Some of his research is cited later on. These publications appear in the middle stages of the concept clusters, but eventually cited. In other words, aspects of his overall research have been necessary for the technologies and in areas in which he was not cited and/or active, he began research that led to it eventually being incorporated and cited.

Summarising the results seen in Figs. 3–5, the scientific publications cited by the patent applications in the INPADOC clusters stem from three sources. These sources include (1) publications cited by the patent applications but not authored by IA, (2) publications authored by IA but not cited by the patent applications, and (3) publications by IA that are cited by the patent applications. The composition of the concept clusters and the period of citing by the INPADOC clusters indicate the relevance at different periods to the technologies of the concepts. The entry of IA publications into these concept clusters indicates a degree of similarity of knowledge and skill sets of IA to the cited publications. The period of entry by IA's publications indicate the adoption of these skills and knowledge by IA. As per the definition of Cohen and Levinthal (1990), absorptive capacity is the ability "... to recognize the value of new, external information, assimilate it, and apply it to commercial ends is critical to its innovative capabilities." In this sense, the entry by IA publications into the NPLR at varying time periods is indicative of the ability of IA to recognise, assimilate and integrate external knowledge into his own skill sets and knowledge base, and these integrations become visible in the scientific background of the patent applications. We also observed IA starting new research in order to improve his absorptive capacity where he seemed to have knowledge and skill gaps.

6. Summary and conclusion

The diverse characteristics of knowledge production, incorporation and dissemination related to product development, lead to a complex model of knowledge capturing. Previous methods used to investigate knowledge transfer have focussed on the facilitating conditions with little emphasis on whether there is any actual knowledge transfer. In this paper, we have explained and developed a method to demarcate and track knowledge transfer. We have done so through combining and modifying existing techniques and the addition of new methodological tools. The resulting method allows one to address the more complex aspects of knowledge capture mechanisms – as illustrated with a stylised start-up or spin-off case.

Through our methods for data processing, clustering and visualisation, we are able to demonstrate the thematic and theoretical links of the inventor's patent output and the inventor's knowledge base and skill sets, as represented by their publication output. This marks a departure from the problems of previous methodologies that relied on individual-specific direct citations to literature by patent applications, in order to determine the theoretical influences of an individual or a field in general (Meyer, 2002).

Through our method, the multi-directional aspects of linkages between science and technology can be examined closely. This allows a quantitative measure of how effectively, and from where exactly, an idea generated in academia is translated into an industrial application or, inversely, how skills and knowledge developed in application may be followed by new research lines generating new scientific knowledge and skills.

Our approach to determining the absorptive capacity of an individual allows us to evaluate the utilisation of scientific knowledge by individuals and their eventual application in technology. The methodology accounts for the influence of co-inventors on the melding of knowledge required for technological output. This allows us to determine the degree and field of contribution from the respective inventors in terms of the base knowledge required for the technology development.

In this study, patent documents and scientific publications are indicators of the content and research behaviours and strategies of the individual researcher/inventor. In its current form, this study does not utilise these indicators to demonstrate the significant anthropological, social and organisational processes inherent in absorptive capacity. Future research tying together the quantitative mapping and tracking of knowledge outputs with qualitative studies examining the more tacit dimensions of knowledge exchange and transfer is necessary to fully understand the emergence and nature of absorptive capacity.

We explained and demonstrated the methodology using a stylised case in which one individual is responsible for much of the growth and success of the start-up firm and bridges both the academic and industrial aspects of knowledge transfer. His research is both fundamental (at his university setting) and applied (in the firms' appropriation and implementation). The researcher–inventor bridges the research environments, facilitating knowledge transfer and skill development between them. In further research we aim to investigate the same processes if there are more star or bridge scientists in one firm, and the effect of their overall contributions.

Our method is not without its shortcomings. Using in-text citations of patent documents requires a significant amount of cleaning due to the differing citation reporting behaviours and requirements across patenting offices. The correct assignment of authors and inventors to publications and patents requires a significant amount of disambiguation. This was handled initially by algorithmic means (Gurney et al., 2012), and then checked manually which was time consuming. The inclusion of all NPLR of the patent applications introduces a level of uncertainty as some of the NPLR are not directly related to the technologies. Many NPLR are very general in nature and address only the fundamental background of the technologies. These NPLR are, without comprehensive expert examination of the patents, difficult to identify but are still included.

This new method of mapping science and technology output and their relations deepens our understanding of the level of contributions of not only individuals and firms, but also of specific institutional policies or models. If a firm or individual performs research within a specific research climate or environment, by utilising the same methodology one would expect to see the overall publishing and patenting activity and the links between the two to vary according to the research climate or environment. In this way, the influence of the environment on knowledge transfer and absorptive capacity can be empirically investigated.

Acknowledgements

This work was supported by a JSPS/CNRS Grant (DP) and AixMarseille University (KS). The authors would like to thank Cristian Martínez, Julie Petidant and Guillermo Lopez Palacios for their work on collecting and processing data. The authors would like to thank the two anonymous reviewers for their valuable contributions to the development of this publication.

References

Arora, A. (1997). Patents, licensing, and market structure in the chemical industry. *Research Policy*, *26*(4), 391–403.
Arundel, A. (2001). The relative effectiveness of patents and secrecy for appropriation. *Research Policy*, *30*(4), 611–624.
Arundel, A., & Kabla, I. (1998). What percentage of innovations are patented? Empirical estimates for European firms. *Research Policy*, *27*(2), 127–141.
Audretsch, D. B., Lehmann, E. E., & Warning, S. (2005). University spillovers and new firm location. *Research Policy*, *34*(7), 1113–1122.
Baba, Y., Shichijo, N., & Sedita, S. R. (2009). How do collaborations with universities affect firms' innovative performance? The role of "Pasteur scientists" in the advanced materials field. *Research Policy*, *38*(5), 756–764.
Bastian, M., Heymann, S., & Jacomy, M. (2009). *Gephi: An open source software for exploring and manipulating networks*. Paper presented at the international AAAI conference on weblogs and social media.
Blondel, V. D., Guillaume, J. L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, *2008*, P10008.
Bozeman, B. (2000). Technology transfer and public policy: A review of research and theory. *Research Policy*, *29*(4–5), 627–655.
Braam, R. R., Moed, H. F., & van Raan, A. F. J. (1991). Mapping of science by combined co-citation and word analysis. I. Structural aspects. *Journal of the American Society for Information Science and Technology*, *42*(4), 233–251.
Branstetter, L. (2005). Exploring the link between academic science and industrial innovation. *Annales d'Economie et de Statistique*, 119–142.
Brouwer, E., & Kleinknecht, A. (1999). Innovative output, and a firm's propensity to patent: An exploration of CIS micro data. *Research Policy*, *28*(6), 615–624.
Callaert, J., Van Looy, B., Verbeek, A., Debackere, K., & Thijs, B. (2006). Traces of prior art: An analysis of non-patent references found in patent documents. *Scientometrics*, *69*(1), 3–20.
Callon, M., Courtial, J. P., & Laville, F. (1991). Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry. *Scientometrics*, *22*(1), 155–205.
Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative Science Quarterly*, *35*(1), 128–152 (Special Issue: Technology, Organizations, and Innovation).
Courtial, J. P., Callon, M., & Sigogneau, A. (1993). The use of patent titles for identifying the topics of invention and forecasting trends. *Scientometrics*, *26*(2), 231–242.
Engelsman, E.C., & van Raan, A. F. J. (1994). A patent-based cartography of technology. *Research Policy*, *23*(1), 1–26.
Fleming, L., & Sorenson, O. (2001). Technology as a complex adaptive system: Evidence from patent data. *Research Policy*, *30*(7), 1019–1039.
Garfield, E., & Welljams-Dorof, A. (1992). Citation data: Their use as quantitative indicators for science and technology evaluation and policy-making. *Science and Public Policy*, *19*, 321–327.
Geuna, A., & Muscio, A. (2009). The governance of university knowledge transfer: A critical review of the literature. *Minerva*, *47*(1), 93–114.
Girvan, M., & Newman, M. E. J. (2002). Community structure in social and biological networks. *Proceedings of the National Academy of Sciences*, *99*(12), 7821–7826.
Gurney, T., Horlings, E., & Van Den Besselaar, P. (2012). Author disambiguation using multi-aspect similarity indicators. *Scientometrics*, *91*(2), 435–449.
Gurney, T., Schoen, A., Horlings, E., Sumikura, K., Laurens, P., Van den Besselaar, P., et al. (2013). *Knowledge capture mechanisms in bioventure corporations: A case study*. Paper presented at the 14th international society of scientometrics and informetrics conference 2013, Vienna.
Hicks, D. (1995). Published papers, tacit competencies and corporate management of the public/private character of knowledge. *Industrial and Corporate Change*, *4*(2), 401.
Horlings, E., & Gurney, T. (2013). Search strategies along the academic lifecycle. *Scientometrics*, *94*, 1137–1160. http://dx.doi.org/10.1007/s11192-012-0789-3
Jansen, J., Van Den Bosch, F. A. J., & Volberda, H. W. (2005). Managing potential and realized absorptive capacity: How do organizational antecedents matter? *Academy of Management Journal*, *48*(6), 999–1015.
Karki, M. M. S. (1997). Patent citation analysis: A policy analysis tool. *World Patent Information*, *19*(4), 269-272.
Lancichinetti, A., & Fortunato, S. (2009). Community detection algorithms: A comparative analysis. *Physical Review E*, *80*(5), 056117.
Langrish, J., Gibbons, M., Evans, W. G., & Jevons, F. R. (1972). *Wealth from knowledge: A study of innovation in industry*. Wiley: Halstead Press Division.
Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. *Research Policy*, *18*(4), 209–223.
Leydesdorff, L. (2008). Patent classifications as indicators of intellectual organization. *Journal of the American Society for Information Science and Technology*, *59*(10), 1582–1597.
Lieberman, M. B. (1987). Patents, learning by doing, and market structure in the chemical processing industries. *International Journal of Industrial Organization*, *5*(3), 257–276.
Magerman, T., Van Looy, B., & Song, X. (2010). Exploring the feasibility and accuracy of latent semantic analysis based text mining techniques to detect similarity between patent documents and scientific publications. *Scientometrics*, *82* http://dx.doi.org/10.1007/s11192-009-0046-6
Martinez, C. (2010). Insight into different types of patent families. In *STI working paper 2010/2*. OECD.
Merton, R. K. (1957). Priorities in scientific discovery: A chapter in the sociology of science. *American Sociological Review*, *22*(6), 635–659.
Meyer, M. (2001). Patent citation analysis in a novel field of technology: An exploration of nano-science and nano-technology. *Scientometrics*, *51*(1), 163–183.
Meyer, M. (2002). Tracing knowledge flows in innovation systems. *Scientometrics*, *54*(2), 193–212.
Mogoutov, A., Cambrosio, A., Keating, P., & Mustar, P. (2008). Biomedical innovation at the laboratory, clinical and commercial interface: A new method for mapping research projects, publications and patents in the field of microarrays. *Journal of Informetrics*, *2*(4), 341–353.
Newman, M. E. J. (2004). Fast algorithm for detecting community structure in networks. *Physical Review E*, *69*(6), 066133.
Nonaka, I. (1994). A dynamic theory of organizational knowledge creation. *Organization Science*, *5*(1), 14–37.
Palla, G., Derényi, I., Farkas, I., & Vicsek, T. (2005). Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, *435*(7043), 814–818.
Pavitt, K. (1998). Uses and abuses of patent statistics. In *Handbook of quantitative studies of science and technology*. North Holland: Amsterdam.
Ponomariov, B., & Boardman, C. (2012). Organizational behavior and human resources management for public to private knowledge transfer: An analytic review of the literature. In *OECD science technology and industry working papers*. OECD Publishing.
Rosenberg, N. (1990). Why do firms do basic research (with their own money)? *Research Policy*, *19*(2), 165–174.
Sagar, A. D., & Van der Zwaan, B. (2006). Technological innovation in the energy sector: R&D, deployment, and learning-by-doing. *Energy Policy*, *34*(17), 2601–2608.
Schmoch, U. (1993). Tracing the knowledge transfer from science to technology as reflected in patent indicators. *Scientometrics*, *26*(1), 193–211.
Schmookler, J. (1966). *Invention and economic growth*. Cambridge, MA: Harvard University Press.
Somers, A., Gurney, T., Horlings, E., & Van den Besselaar, P. (2009). *Science Assessment Integrated Network Toolkit (SAINT): A scientometric toolbox for analyzing knowledge dynamics*. The Hague: Rathenau Institute.
Tamada, S., Naito, Y., Kodama, F., Gemba, K., & Suzuki, J. (2006). Significant difference of dependence upon scientific knowledge among different technologies. *Scientometrics*, *68*(2), 289–302.
Tijssen, R. J. W. (2001). Global and domestic utilization of industrial relevant science: Patent citation analysis of science–technology interactions and knowledge flows. *Research Policy*, *30*(1), 35–54.

Tijssen, R. J. W. (2006). Universities and industrially relevant science: Towards measurement models and indicators of entrepreneurial orientation. *Research Policy*, *35*(10), 1569–1585.
Tijssen, R. J. W., & Van Raan, A. F. J. (1994). Mapping changes in science and technology. *Evaluation Review*, *18*(1), 98–115.
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).
White, H. D., & McCain, K. W. (1998). Visualizing a discipline: An author co-citation analysis of information science 1972–1995. *Journal of the American Society for Information Science and Technology*, *49*(4), 327–355.
Whittaker, J. (1989). Creativity and conformity in science: Titles, keywords and co-word analysis. *Social Studies of Science*, *19*(3), 473.