DRUID

Paper to be presented at DRUID18
Copenhagen Business School, Copenhagen, Denmark
June 11-13, 2018

The relation between scientific and technological knowledge in emerging fields: Evidence from DNA Nanoscience and DNA Nanotechnology

Hanh Luong La
Eindhoven University of Technology
School of Innovation Sciences
l.h.la@tue.nl

Rudi R.N.A. Bekkers
Eindhoven University of Technology
School of Innovation Sciences
R.N.A.Bekkers@tue.nl

Abstract

To better understand the relationship and interaction between scientific knowledge and technological knowledge, previous studies mainly focused on observing similarities between these two domains. This paper aims to extend our understanding of this relationship by first looking at the interaction between these domains based on the stylised views of science-push and technology-pull, and then detect their similarities, differences, and potential complementarities based on the evolution of their knowledge content. We employed a novel method, called the concept approach for selecting publications and patents in an emerging knowledge field, namely DNA Nanoscience/Nanotechnology, then performed the co-occurrence of terms. Our results not only show that both elements of the stylised 'science push' and 'technology pull' views can be recognised but also suggest that co-evolution plays a much more prominent role.

The relation between scientific and technological knowledge in emerging fields: Evidence from DNA Nanoscience and DNA Nanotechnology[1]

Hanh La, MPhil.[†*], Prof.dr.ir. Rudi Bekkers[†]

[†]School of Innovation Sciences, Eindhoven University of Technology, The Netherlands

*Contact: L.H.La@tue.nl

Abstract

To better understand the relationship and interaction between scientific knowledge and technological knowledge, previous studies mainly focused on observing similarities between these two domains. This paper aims to extend our understanding of this relationship by first looking at the interaction between these domains based on the stylised views of science-push and technology-pull, and then detect their similarities, differences, and potential complementarities based on the evolution of their knowledge content. We employed a novel method, called the concept approach for selecting publications and patents in an emerging knowledge field, namely DNA Nanoscience/Nanotechnology, then performed the co-occurrence of terms. Our results not only show that both elements of the stylised 'science push' and 'technology pull' view can be recognised, but also suggest that co-evolution plays a much more prominent role.

Keywords: knowledge similarity, knowledge complementarity, science and technology (S&T), concept approach, text-mining

[1] Paper accepted for presentation at the conference DRUID18, Copenhagen, 11-13 June 2018.

1. Introduction

The relationship between science and technology (S&T), as the two principal knowledge domains, has generated numerous discussions among innovation scholars, philosophers, and historians of technology. It is generally understood that S&T are two interacting and interdependent entities (Breschi & Catalini, 2010; Meyer, 2000), and also co-evolve (Murray, 2002). However, few empirical studies offer evidence on the nature of this relationship (Han and Magee, 2017), and even less so in emerging technology fields. A better understanding of this relation in emerging technology fields is desirable, as these are the fields that are expected to provide promising solutions to societal challenges, as well as to economic growth.

While government funding into nanotechnology generally remains substantial since the mid-1990s, there is a growing need for evaluating its impact on academic publications and patents (Tahmooresnejad & Beaudry, 2015). From a policy perspective, it is particularly interesting which types of intervention and funding have produced the highest impact on the development of emerging fields. Is it better to fund basic and applied research, so that it would then later automatically lead to innovative technologies? Or is it better to fund technological developments, so that they would then stimulate the basic research?

One of the central thoughts in European policy is known as the 'European Paradox', which implies a gap between the generation of high quality scientific output and the creation of technological innovations (see European Commission (1995). Similar observations were found in Zhao & Guan (2013) for nanotechnology in some leading universities (although the thought of the European Paradox is disputed by others, see for instance Dosi et al., 2006). It is worth mentioning that nanotechnology embodies a wide range of branches and applications. Some of these were successfully commercialised into marketable products and services. However, some others have ended at the proof of concept period. Therefore, further understanding of the S&T relationship in emerging technologies, such as nanotechnology, can also help in informing decision-makers and policy discussions and interventions, aiming at higer economic growth or sustainability. At the same time, research into such fields is challenging because their body of knowledge is not yet well-established, but fast-moving with interesting potential trends.

This paper aims to understand the interaction and relation between scientific and technological knowledge of the emerging field of DNA Nanoscience/Nanotechnology (DNA-NT), one of the most promising branches of Nanotechnology. This field has shown significant advances in scientific knowledge and technological knowledge and is promising regarding a wide range of application areas, including health (structural biology, biophysics, and nanomedicine), but also in other areas like molecular scale electronics.

The paper is structured as follows. Section 2 discusses the current literature and research questions. Section 3 introduces our research methodology, including a new method of building databases of publications and patents in emerging, challenging fields, called the concept approach. Section 4 provides an overview of our actual data, and Section 5 presents our empirical results. Section 6 summarises our findings, offering research contributions, and research implications. Section 7 confers our limitations and future research.

2. Theoretical background: the inter-relationship between scientific and technological knowledge

The relationship between scientific and technological knowledge is widely discussed in a larger interesting theme: the relationship between S&T. Most of the empirical literature discussing S&T relationship and interaction seeks to better understand their overlap. In this paper, our intention is not to focus on the complexity of integration, but rather breaking both domains down (or decomposing them) into core knowledge areas - while examining their interaction and comparing their knowledge content and structure. To investigate the time patterns of S&T interaction, we would like to take science-push and technology-pull, two stylised views in the S&T literature. To compare their knowledge content, we not only look at their similarities, but also their differences, and potential complementarities.

Section 2.1 discusses theoretical contributions regarding the direction of influence between the two domains, especially focusing on the stylised views of science-push and technology-pull. Section 2.2 reviews existing studies that investigate similarity and complementarity, the two most observable patterns of knowledge in the domains of S&T. We concluded both sub-sections by raising our research questions.

2.1. The time dimension of S&T interaction: Science-push and Technology-pull

A huge body of literature has considered the relation between S&T in terms of their mutual influence and proposed stylised views of science-push and technology-pull theories. This section will discuss these two views.

Most of the **science-push** papers claim that technology is dependent on science. In early days, Vannevar Bush (1945) proposed the linear model that argues that 'science pushes technology', and in turn, technology (especially R&D) leads to production and diffusion. In this view, science always precedes technology. Although this early innovation model is often criticised as too simplistic and unrealistic (Edquist & Hommen, 1999; Mowery & Rosenberg, 1979), more recent empirical evidence, yet emphasises the role of science and universities in creating 'technological opportunities' (Klevorick, Levin, Nelson, & Winter, 1995). Being influenced by the linear model, another common view that 'technology is applied science' was also popularised in the literature (Audretsch et al., 2002). In that view, scientific knowledge must emerge before technological knowledge can further develop. This single one-way view opened a debate in science education, as discussed by Gil-Pérez et al. (2005). By using a historical view, this author supported the claim by Gardner (1994) that technology has preceded science by thousands of years. And most science teachers at that moment were not aware of this. Narin's quantitative approach (1985; 1997) of studying paper citations in patents has been quite influential in examining the 'science-push' effect towards technology. However, according to Meyer (2000), from such citation analysis, it would be insufficient to conclude that there is a direct link between a focal patent and its cited papers. The results of Meyer's in-depth interviews with inventors of 10 patents show that in some cases, cited scientific publications, although being important, only provides background information for the invention. By analysing the S&T interaction related to the most active 20 universities in nanotechnology, Zhao and Guan (2013) found that in nanotechnology the effect of 'science-push' is stronger than 'technology-pull'. They claimed that in those universities, because nanotechnology is still in the early phase of development, technology cannot precede and lead scientific progress effectively.

The view in the literature known as **technology-pull** adopts quite the opposite view. Rosenberg (1982) stated that scientific progress increasingly depends on technological and economic inputs. It is an endogenous activity, rather than an exogenous one. When it comes to specific technological inputs, several scholars (among whom, Layton, 1974; Rosenberg 1992) have emphasised the role of instrumentation (besides scientific theories and practice) in developing scientific outputs. In Zucker & Darby (2005), technology is seen as a crucial scientific instrument, or the 'invention of the method of inventing,' which embarks on new scientific transformations. Notably, in nanotechnology, the role of advanced microscopes, such as the scanning tunnelling microscope (STM) and the atomic force microscope (AFM) is indispensable in approaching and designing nanoscale objects (Darby & Zucker, 2003; Zucker & Darby, 2005). In addition to the context of nanotechnology, it is interesting to mention that it may be related to more applied sciences, rather than basic ones. In basic sciences, natural phenomena are the objects of inquiry. By contrast, in 'sciences of the artificial' (Boon, 2009; Perkmann, 2009; Simon, 1969), phenomena in technological artefacts are researched. These kinds of sciences are aimed at providing knowledge that is employed further in technological design (Boon, 2009). As far as 'science of the artificial' are concerned, technology might come earlier than science, and then becomes the object of scientific research. In its turn, science supports further technological improvement. Along the same line as technology-pull, Bernardes and Albuquerque (2003) introduce the 'inverted linear model', in which economic development initiates technological development, and then the development of scientific institutions. This model is discussed in the context of less developed countries, where more resources are necessary for the development of S&T. However, Bernardes and Albuquerque (2003) also raised a question of the source of economic development in those countries, if there are no investments in S&T. Therefore, the 'inverted linear model' may not be the right substitution for the traditional linear model.

It should be noted that there is also the concept of science-push vs. market-pull, where S&T are on the same side, and market is on the side. The work of Schmoch (2007) suggests double-boom cycles which include a science- and technology- push, and a market-pull, wheras Yu & Yingming (2008) represent a reverse double-boom cycle evolutionary model, in which a technological trajectory moves from a 'market-pull' to a 'technology-push' cycle, mostly focused on the context of less-developed countries.

The stylised models of science-push and technology-pull also triggered a discussion on how linear (or inverted linear) these developments really are. In an evolutionary vein, scholars have proposed a more 'non-linear' thinking, in which both the effects of 'science-push' and 'technology-pull' are present together with many other interactions in a national innovation system (Nelson and Rosenberg, 1993).

With exloring the time dimension of the S&T interaction, we would like to raise our first research question:

- *To what extent does knowledge development follow the stylized 'science push' and 'technology pull' patterns?*

2.2. Comparing knowledge development in science and in technology

In this section, we review the literature on comparison between S&T as two knowledge domains and the literature on comparison between sub-domains in the same domain (Science or Technology), with the purpose to find directions (or shed light on) for our methodology and empirical analysis. Table 1 shows us some typical literature on the comparison of knowledge development between S&T. Basically, there are two main streams: Both streams admit the interaction between S&T, however, the first stream (Quadrant I and IV) considers S&T as two separate entities, the second stream (Quadrant II and III) considers studies them as two integrating entities. Our study should be put on the Quadrant IV. It is also a comparative study, which is based on both theoretical and empirical literature[2], with more emphasis on the empirical side.

Table 1. Some quadrants of research on S&T relationship with examples

	First stream of literature considers S&T as two separate entities	**Second stream of literature mainly looks at the overlap between S&T**
Theoretical	**<I>** Price (1965), Dosi (1982), Pavitt (1987)	**<II>** Layton (1974), Nordmann (2008), Arthur (2009)
Empirical	**<IV>** Mina (2007), Zhao and Guan (2013), ***and our study***	**<III>** Murray (2002), Breschi and Catalini (2010)

In Quadrant I and IV, it is apparently distinguishable between two streams. It is evident from *Quadrant I* that S&T are characterised by similarities and complementarities. Examples of the scholars who followed this approach are Price (1965), Dosi (1982) and Pavitt (1987). The work by Price (1965) is considered as one of the earliest important studies on the co-evolution that we are interested in. Price referred to the Toynbee's 'pair of dancers' as a metaphor for this relationship. The work of Price (1965) implies that S&T are two (parallel) co-evolving cumulative autonomous structures/entities. Although the dancers could be male and female, having differences in attitude and structure, they move to the **same** music. These dancers of S&T in Price (1965) are characterised by 'infrequent interaction,' 'separate cumulating structure,' and more interestingly, are considered to be **complementary**. Two decades later, Dosi (1982) described the domains in terms of scientific and technological paradigms, and of scientific and technological trajectories. He reiterated Thomas Kuhns (1970) view of a scientific paradigm as a model, a pattern, a set of problems of inquiry. In analogy with Kuhn's scientific paradigm, Dosi defined technological paradigm as a 'model, a pattern of solution of selected technological problems, based on selected principles derived from natural sciences and on selected material technologies'. In this sense, the **similarities** between scientific and technological paradigms lie in the mechanism and procedure of both S&T. Pavitt (1987) strongly argued that the efficiency of the whole field is not inevitably an outcome of creating more similarities between S&T. He emphasised that policy-making should consider the **complementarity** between S&T, which

[2] It does not only investigate similarities, but also complementarities and differences, as being mentioned earlier.

'varies considerably amongst sectors of application, in terms of the direct usefulness of academic research results, and of the relative importance attached to such results and to training'.

Quadrant IV includes empirical studies that view S&T as two separate knowledge domains. This part of literature is understudied. Mina et al., (2007) compared the evolution between scientific and technological knowledge of treatment for coronary artery disease by comparing two main paths[3] of its scientific and technological networks. Nevertheless, in terms of network analysis, their methods were based on a limited number of the most highly weighted links on both networks, rather than stressing on the importance of the nodes (which are publications and patents). From a different perspective, Zhao & Guan (2013) introduced a new model characterising the relationship between S&T based on their own classification of S&T styles and the changes in production of publications and patents. Their model is distinctive from the rest of the literature, however, their datasets are only limited to the top 20 universities in nanotechnology. Therefore, the role of industry in publishing and patenting was ignored.

Quadrant II investigates the S&T knowledge relationship via the integration or the overlap between S&T. However, while two domains are observed from their overlap, it is easier to find similarities, rather than complementarities[4]. Accordingly, the differences between S&T turn out to be smaller (Meyer, 2000). In a theoretical contribution, Layton (1974) explained the transformation of a set of technological rules into a new entity of science: 'technological science' or 'engineering science.' In a similar vein, Arthur (2009, p. 61) further articulated that S&T are 'deeply woven' into each other. In fact, in the field of nanoscience and nanotechnology, a few scholars articulated the term 'nanotechnoscience' (Nordmann, 2008; Patra, 2011). Such terms refer to the integration or merger between science and technology, a context in it is more difficult to distinguish between S&T, or between basic and applied research (Nordmann, 2008).

In Quadrant III, the relationship between S&T has been mostly investigated via the overlap of scientific and technological networks. Murray (2002) examined the co-evolution of S&T via patent-paper pairs, based on the assumption that a single idea is described in both a patent and a paper. From such pairs, networks of co-authoring and co-patenting were established for further analysis. Murray's work was the foundation for Breschi and Catalini (2010), who traced the link between scientific and technological networks via their gatekeepers: the key actors who engaged in both publishing and patenting. Perhaps, however, the emerging topics around these gatekeepers are just the tip of the iceberg, which reflects just the part of both networks where the similarities are the strongest. This ignores the overall knowledge content of both S&T, which was created by non-author inventors and non-inventor authors. If this is true, then it would be good to look at the S&T relationship also from a broader view, i.e. various patterns of interaction and of co-evolution, rather than only based on similarities as shown by patent-paper pairs.

In summary, theoretical literature on S&T relation is characterised with similarities and complementarities. Empirically, however, S&T relation is investigated mainly via similarities and barely via differences, but not complementarities.

Related back to our study, we are interested in comparing the dynamics of knowledge content in both domains over time. We then came up with our second research question:

[3] Representing the most important developments in the citation networks of publications and patents

[4] This becomes true with the empirical literature where scholars investigate the overlap between scientific and technological networks.

- *To what extent the knowledge content in these domains are either similar, complementary, or different in their evolution process?*

When we discuss the knowledge content, we consider S&T as two text-based knowledge domains, where we can find single occurrences and co-occurrences of different pieces of knowledge (knowledge sub-domains/knowledge areas).

3. Research methodology

3.1. Creating a dataset with high recall and precision: the concept approach

In order to investigate the evolution of knowledge in patents and knowledge in publications, we first need to select patents and publications and put them in two sets. The set of patents represent the technological domain, the set of publications represent the scientific domain. Next, we operationalised each domain (i.e. broke down two sets) into smaller units: knowledge areas, which are represented by the most significant terms in publications and patents. The key of building a dataset of publications and patents is a good selection procedure, and a good dataset is characterised by having both a high degree of recall and precision (Xie & Miyazaki, 2013). Recall is defined as the proportion of relevant records retrieved. Precision is the proportion of retrieved records that are relevant. Ideally, recall and precision are both perfect, but in real life this is usually not possible, and there is always a trade-off between the two: one can easily achieve perfect recall but at the expense of lower precision, or the other way around. The challenge is to find an appropriate balance between recall and precision.

The achievable levels of recall and precision also depend on the subject area and the novelty of the field. In emerging technologies, tracking patents and publications is often challenging, because of the high possibility of having poorly defined data (Huang, Schuehle, Porter, & Youtie, 2015). There are no established journals or patent classes for these types of emerging fields. The challenges that the researchers often face are either low recall or low precision, or the imbalance among the sub-areas of the emerging field (*ibid.*). Whereas Porter et al. (2008) emphasise that for a vast domain like nanotechnology, there is no absolute standard for recall and precision, Huang et al. (2015) suggest that a search with high recall and 'satisfactory precision' (>50%) is useful in emerging technology studies. In our study on DNA-NT, we follow the latter approach: we aim to balance recall and precision but with an emphasis on recall over precision.

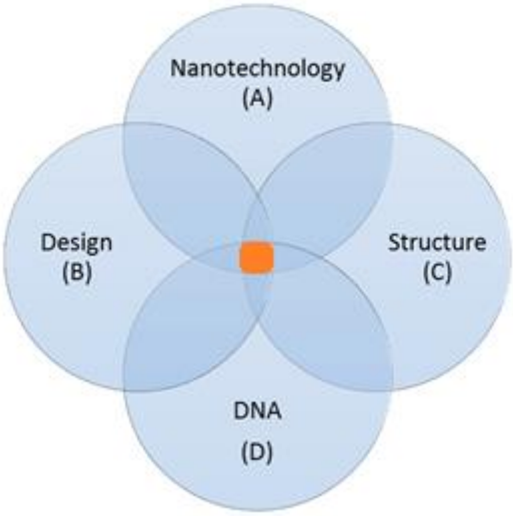
For both publications and patents, the most common search/selection strategies are keyword search and classification search (Benson & Magee, 2013), or the combination of the above. The keyword search uses search terms in combination with Boolean operators. The classification search is applicable when publications are classified in research areas (e.g. Web of Science categories), or where patents are classified in classes/sub-classes (e.g. IPC codes), or application areas. More sophisticated approaches for keyword search are those using structured text-mining software and expert inputs to identify key terms (see Arora et al, 2013). Other approaches for classification search include the Classification Overlap Method, which splits the definition of a technology into two components, a functional or 'artefact' component and a 'knowledge' one (Benson & Magee, 2015).

Neither keyword search or classification search strategies are perfect, however. Some testing we did in our own area of interest, DNA-NT, showed poor levels of recall and precision, most likely because this is an emerging complex technology field, whose boundaries with other knowledge fields (e.g. bionanotechnology) are fuzzy and developing. Keywords that distinguish this field from adjacent fields are hard to find. Classifications are rare or do not exist for emerging fields. In most cases, they are generally non-compatible when making comparisons between publications and patents.

Eventually, we adopted an approach which we learned from intensive interaction with technology and business intelligence units from industry that work on patent landscaping and other patent studies. To the best of our knowledge, this approach is new in scholarly studies, and we will refer to it as the 'concept approach'. In short, it works as follows: First, one operationalises the definition of a *knowledge domain* into several concepts that each represent an indispensable element of the field in question. In each concept, one tries to find all alternative words (e.g. synonyms) related to the concept. The concepts result in an as complete selection as possible, i.e. many publications / many patents (so these concepts themselves have a high recall but low precision). As a second step, one selects the intersection of all concept groups and ideally ends with a much smaller set, with exactly the publications and patents one aims to find (high recall *and* high precision).

In our study, we focus on the knowledge field of 'DNA Nanotechnology' (terminology that is often used in both publications and patents), and 'DNA Nanoscience', by which we here mean the scientific domain of DNA Nanotechnology (see Douglas, 2016 for a more elaborate discussion on the concept of DNA Nanoscience). On the website of Nature, DNA Nanotechnology is defined as 'a branch of nanotechnology concerned with the design, study and application of synthetic structures based on DNA. DNA Nanotechnology takes advantage of the physical and chemical properties of DNA rather than the genetic information it carries'. Based on our representative literature review and experience in interacting with professionals in DNA-NT, we derived four related sub-concepts to be used in our concept approach: **Nanotechnology** (A), **Design** (B), **Structure** (C) and **DNA** (D), as illustrated in Figure 1. Under each sub-concept area, we maximised our search by adding more synonyms or relevant keywords, which were collected exhaustively from all possible sources. Those information sources include the above Nature's definition, the materials and notes taken by this researcher from conferences, the communication with experts via Skype and emails, the publications and news in the field that this researcher is familiar with. Subsequently, we selected the overlap among those four above concepts. To improve the precision of the dataset, we introduced some exclusion terms to remove the irrelevant records from the datasets. After that, we sent out our queries to experts for validation: one expert from industry, and the other from university. Then, we revised our queries for the last time and started to collect the data. Because the patent dataset is often much smaller than the publication dataset, we complemented the found patent data with other patents that cite the patents in the set we found or are cited themselves by the patents in the found set. This step further increases recall, without a notable drop in precision. Annex 1 provides details on the concepts used, and the final search queries we used.[a]

Figure 1. Illustration of the concept approach for DNA Nanotechnology



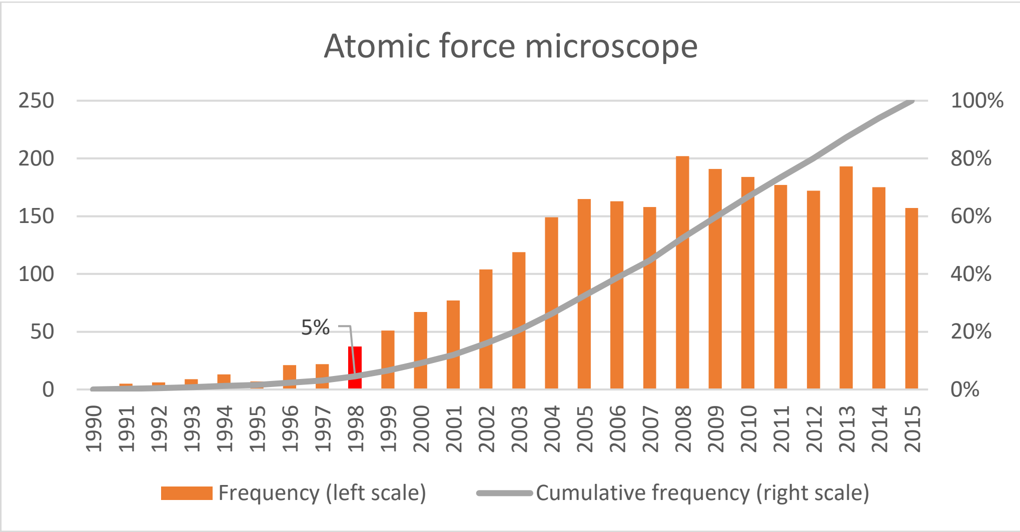
3.2. Methods

In a nutshell, to investigate the time dimension of S&T, and comparing these two knowledge domains, we performed text-mining of two unstructured sets of documents (a set of publications and a set of patents), transformed them into structured comparable data (representing by the most significant knowledge areas/terms), then compare their content development over time.

Our first task is to test the time dimension of 'science push' and 'technology pull' in this specific knowledge field of DNA-NT. We test this by taking narrowly defined knowledge areas ('technical terms') that are present in both domains, and investigate differences in time patterns (Schmoch, 2007). For a given technical term, we will consider it to be a case of science-push when the term appears in the science domain at least five years before it appears in the technology domain. The other way around, we will consider a particular term to be a case of technology-pull when the term appears in the technology domain at least five years before it appears in the science domain. Time gaps that are smaller, or non-existent, in contrast, suggest co-evolution of knowledge, instead of science-push / technology-pull.

To determine the exact precise moment that a term first appears in the science or the technology domain, we consider the year of publication or the year of patent filing. However, we want to prevent that our determination of this moment is merely driven by a chance occurrence of a term only. Instead, we want a certain critical mass to be achieved reflecting that knowledge has started to develop in the domain in question. Therefore, we applied a threshold: we consider the appearance of a term to be the moment when that term hits the 5% threshold of its cumulative frequency over all periods. For most of our terms, this threshold is met at the absolute value of approx. 100 documents or higher. Figure 2 presents an example of the threshold we used. In the scientific domain, the term 'atomic force microscope' is first mentioned in the year 1991. In the following years, its occurrence grows a bit, by 1995 it drops a bit, then starts growing again. By the year 1998, it reaches our 5% threshold: at that time, it has been mentioned 120 times, out of a total of 2624 times at the end of our investigated period.

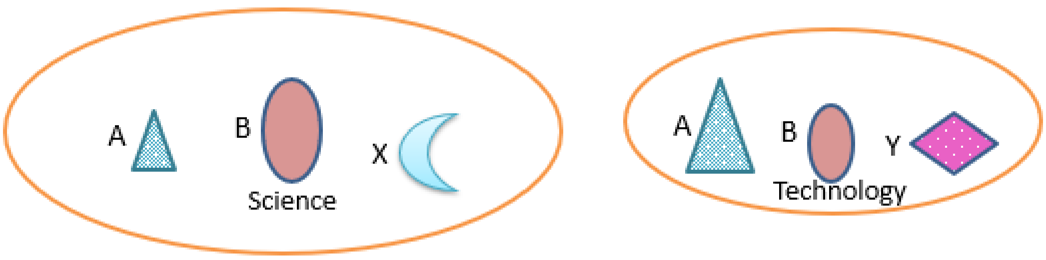
Figure 2. Illustration of the occurrence of the term 'Atomic force microscope' in the scientific domain



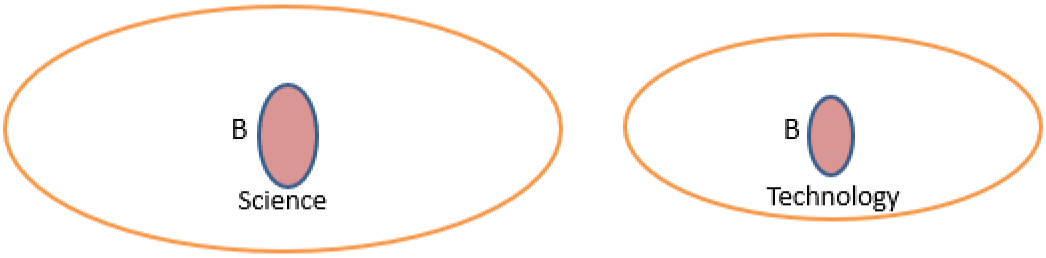
Our second task is to define and measure knowledge similarity and complementarity between scientific and technological knowledge, and identify the specific knowledge areas where the phenomenon of similarity and complementarity occurs. During the search for knowledge similarity and complementarity, we also need to tackle knowledge commonality and difference. We discuss all cases of knowledge relatedness starting from commonality to difference. In those series, *knowledge commonality* occurs when the same narrowly defined knowledge area appears in both domains, regardless of its extent/frequency in each domain. *Knowledge similarity* itself is a stricter concept than *knowledge complementarity*. It occurs when the same narrowly defined knowledge area appears in both domains with the same relative extent. It may imply that in a larger domain, the actual appearance of the focal narrowly defined knowledge area can be higher than its actual appearance in the smaller domain (Figure 3-b). *Direct knowledge complementarity* happens when two distinctive narrowly defined knowledge areas are strongly adjacent in terms of knowledge mapping. In addition, we also considered *indirect knowledge complementarity* between two knowledge areas happens when each of them is strongly adjacent to a third knowledge area, which appears in both domains. Figure 3 illustrates the mentioned above knowledge relatedness concepts.

Figure 3. Different levels of knowledge relatedness: Commonality, Similarity, Complementarity (Direct and Indirect) and Difference.

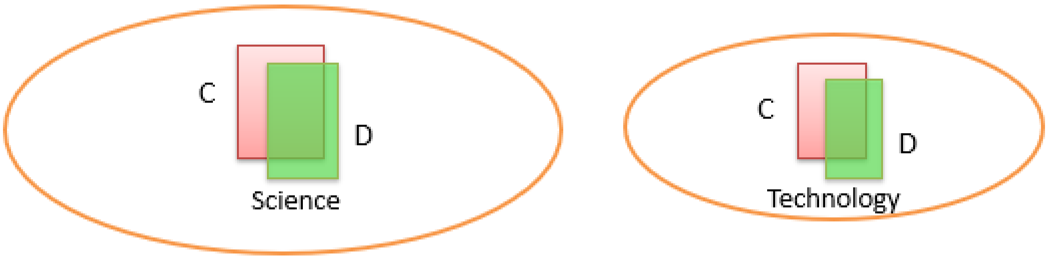
a. ***Common and different*** *knowledge areas*: A and B (separately) are *common* knowledge areas because they appear in both S&T. The relative extent of a common knowledge area can be different (A) or similar (B) in each domain. X and Y are *different* knowledge areas because each of them only appears in one domain.



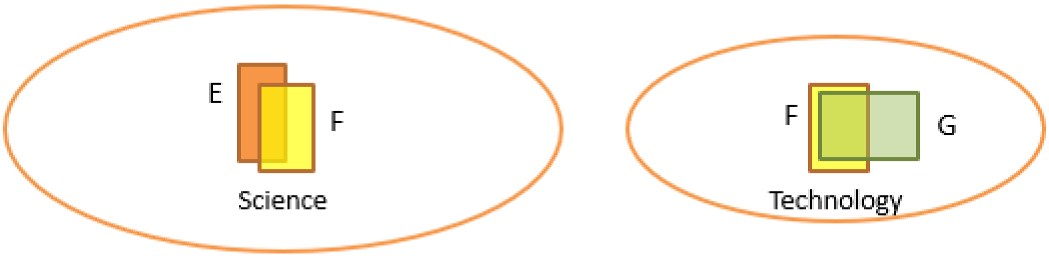
b. ***Similar** knowledge areas*: B is a similar knowledge area because it appears in both domains with similar relative extent. Similar knowledge areas are *per definition* common ones, but not the other way around.



c. ***Directly complementary*** knowledge areas: X and Y are complementary knowledge areas when they strongly co-occur in both domains. Their extents of usage overlap in a significant number of documents. In terms of knowledge mapping, they are adjacent.



d. ***Indirectly complementary*** knowledge areas: X and Z are complementary knowledge areas when one finds another knowledge area Y being strongly adjacent to X in Science and being strongly adjacent to Z in Technology.



The above definitions were articulated at individual term level. To compare two domains, we need to aggregate our results to the domain level. Regarding *knowledge commonality*, we tried to identify all narrowly defined knowledge areas (represented by terms) that two domains have in common in different (sub-) periods, regardless their extent. To measure *knowledge similarity*, we aimed to check if those common knowledge areas appear to the same extent in both domains. From the list of common terms, we performed the Chi-square test for corpus similarity to assess whether both domains consist of terms drawn randomly from some larger population (for this test, see Kilgarriff, 2001 and Evert, 2005). Therefore, we considered the corpora to be similar (i.e. belonging to some larger population) in respect to each term when the outcome is significant (i.e. the p-value of the test in respect to each term is less than 5%).We repeated the Chi-square test not only for the full period, but also separately for three sub-periods. In each (sub-)period, we found a list of similar terms, i.e. terms appearing with similar relative frequencies.

Among the remaining terms appearing with different frequencies, we wanted to find out pairs of complementary terms. As we know of no existing cross-domain measure of *complementarity*, we came up with our own tests, which can be applicable to any two knowledge domains. Our analysis of complementarity derives from the co-occurrences of terms.

The first test measures the direct complementarity between two knowledge areas (represented by two terms). Direct complementarity between X and Y happens when two terms strongly co-occur in both domains.

In Science: X_n and Y_m co-occur with Jaccard index J_i

In Technology: X_n and Y_m co-occur with Jaccard index Jj

The joint Jaccard index of X_n and Y_m co-occur in both S&T is calculated as follows:

$$J_{direct} = \sqrt{J_i \times J_j}$$

When the Jaccard index in each domain is high, the joint Jaccard index is also high. Then we can conclude that two terms strongly co-occur in both domains, or they are complementary knowledge areas.

The second test measures the indirect complementarity between two knowledge areas. Indirect complementarity between two terms X and Z happens when one finds another third term Y strongly co-occurring with X in Science and with Z in Technology.

In Science: X_m and Y_i co-occur with Jaccard index J_{im}

In Technology: Z_n and Y_i co-occur with Jaccard index J_{jn}

The joint Jaccard index between X_n and Z_m is calculated as follows:

$$J_{indirect} = \sqrt{J_{im} \times J_{jn}}$$

When the Jaccard index in each domain is high, the joint Jaccard index is also high. Then we can conclude that two terms X and Z are complementary because they strongly co-occur with the same term Y in both domains.

Empirically, for both cases of knowledge complementarity, we followed two steps: First, we reduced the co-occurrence networks of 110 terms to smaller networks with only edges with Jaccard index larger than 0.01[5]. Second, we matched common pairs between S&T, calculated the Joint Jaccard index[6], sorted and compared the lists of direct and indirect complementarity.

4. Data

4.1. Publication and patent datasets

Following the concept approach and other steps introduced above, using Web of Science (to retrieve publication content), in Derwent Innovation Index and PATSTAT version Autumn 2016 (to retrieve patent content). We found 135,055 publications and 11,226 patents, dated between 1947 and 2015. However, because the data on academic publications prior to 1990 often lack abstracts, we truncated our data to the period between 1990 and 2015. We cleaned duplicates and documents without titles and got two datasets: one contains 123,929 publications, the other contains 10,476 patents. To allow for sub-period analysis, we further divided this 26 years of

[5] This first step resulted in 538 pairs in Science and 391 pairs in Technology.

[6] This second step resulted in 133 pairs of direct complementarity and 10525 pairs of indirect complementarity.

data into three sub-periods: Sub-period 1: 1990-1997; Sub-period 2: 1998-2005; Sub-period 3: 2006-2015. The break point between Sub-period 2 and 3 is based on a ground-breaking contribution by a Caltech researcher Paul Rothemund, published in 2006 in Nature (receiving over 2,500 citations) (Rothemund, 2006) and patented in 2007. Before that, no such compelling breaking point existed, so we chose to give the two previous periods an equal length.

4.2. Data for knowledge content

To obtain empirical data for the analysis of knowledge content, we need to extract the most significant terms with accurate frequencies from the datasets of publications and patents. Each term should reflect well the knowledge content in each knowledge area, i.e. capture the most important concept in the domain. We think that title and abstract are the most important representation of the knowledge content of a document. Titles contain the most carefully selected words by the authors or inventors, however, offer a limited number of concepts. Abstracts contain sufficiently most important concepts and briefly explain the contribution of authors and inventors. Using both titles and abstracts as the document representation is a more efficient way to save our time and avoid computation and filtering problems.

When it came to extracting the core technical terms from the two corpora containing titles and abstracts of publications and patents related to DNA-NT, we had to deal with several challenges:

(1) Technical terms often exist of a combination of words, not a single word (Nakagawa, 2000). The field we study is not an exception. Single words appearing with high frequencies[7] (e.g. DNA, temperature…) are often not sufficiently descriptive regarding the main contributions of authors. Using a high Term Frequency- Inverse Document Frequency (tf*idf) does not solve this problem. High-frequency terms can become meaningful, descriptive terms, however, if they are combined with other words into compound nouns (e.g. DNA origami, temperature control…). We addressed this challenge by using the automatic Term Recognition algorithm described in Nakagawa (2000) and integrated as Term Extract function in KH coder. In short, a Term Extract score is computed on the basis of how many compound nouns have a simple noun N included as an element. In other words, the more frequently a simple noun is integrated to other compound nouns, the higher score it gets.

(2) Frequently occurring compound nouns can still be non-technical or non-descriptive, or may fall outside our knowledge area of interest. As no software or algorithm can solve this in a fully automated way, we addressed this challenge by extensive manual checking and exclusion. As part of this manual checking, we excluded POS (Part of Speech) words and other generic biological terms such as DNA, RNA, protein, acid amine, etc.

(3) Certain terms can be written in more than one way. Techniques such as stemming (i.e. cutting ends of words, for instance from ‘saying’ to ‘say’) or lemmatization (i.e. finding the original form of a word, for instance from ‘said’ to ‘say’) may be helpful for some words, especially verbs, but will not work for other words, and not for synonyms and abbreviations. To address this challenge, we manually harmonised terms (such as grouping synonyms, abbreviations) into term groups (or codes). For example, we

[7] And even with high Term frequency- Inverse document frequency (tf*idf)

harmonized ‘3D structure’ into ‘three-dimensional structure’, ‘control of temperature’ into ‘temperature control’, and ‘au nanoparticle’ into ‘gold nanoparticle’.

After the above processes, we found 110 harmonised terms that appear solely or simultaneously in our two corpora. Finally, we used KH coder (see above) to count the document frequency (the number of documents where a term occurs at least one time) of those terms in our datasets across years and periods.

5. Empirical analysis and results

5.1 Science-push vs. technology-pull

Using the publication and patent datasets, we investigated timing patterns, using the methodology introduced above. From the original 110 terms in the data set, we excluded 19 terms that appear in only one domain, or that have frequencies in at least one of the domains of less than five, leaving us with 91 terms for which we measure time differences.

As explained in the methodology, we looked for cases where the time difference of appearance (using a threshold of 5%) exceeds five years. For 18 of the total of 91 terms, this is indeed the case. Among those 18 terms, seven terms represent strong signals for ‘science-push’, 11 terms represent strong signals for ‘technology-pull’. Regarding the remaining 73 terms, we did not observe such a long time gap between S&T. Therefore, we can consider these 73 terms as signals for the co-evolution of those knowledge areas between the two domains. Below, we discuss the science push, technology pull and (potential) coevolution cases in more detail.

Table 3 shows the list of 7 **science-push** terms, with the longest time lag being observed in ‘x-ray crystallography’ and ‘crystal structure’. Among other terms, we found the input (Ecoli), output (DNA structure), techniques (Raman spectroscopy) and the aim for ‘high stability’ that is set by early scientists in the field. In these areas, science proved itself as a strong foundation for DNA-NT. It had reached its critical mass much earlier before the development of technology.

Table 3. List of terms with strong signals for science-push

	Term	Threshold reached of 5% total frequency of each term in publications (1)	Threshold reached of 5% total frequency of each term in patents (2)	Time gap between (1) and (2)
1	X-ray crystallography	1993	2004	-11
2	Crystal structure	1994	2002	-8
3	Ecoli	1992	1999	-7
4	Raman spectroscopy	1994	2001	-7
5	High stability	1995	2002	-7
6	DNA structure	1993	1999	-6
7	Molecular biology	1992	1997	-5

Since the early period, science was the one who pushed technology in building up DNA nanostructures, using the existing techniques, instruments and inputs in previously existing fields, namely molecular biology and crystallography. Regarding X-ray crystallography, one of the existing techniques, our result is in line with Pinheiro et al. (2011): before founding the field, Nadrian Seeman, a crystallographer, had worked on organising proteins into 3D crystals, so that he could study their structure with X-ray crystallography. Regarding instruments, raman spectroscopy is quite a favorite technique for identifying molecules and studying chemical bonding, which is extremely helpful in creating artificial DNA nanostructures. Regarding inputs, DNA-NT might have used Ecoli, a popular input in molecular biology. Furthermore, scientists at that time started to be concerned about whether those structures have ‘high stability’, and how to make it happen.

Among the 7 science-push terms, we have seen the names of two technologies (X-ray chrystallography and raman spectroscopy). It may be implied that these technologies are science-enabling tools. During the interaction process between S&T, science led technology in terms of publications using the scientific instruments offered by technology. Therefore, we can clearly see that the science-push effect should not be seen as purely one-way influence, but involves with more complicated processes. However, the interaction occurs in the labs, even in the same researchers.

The persistent development and influence of science have resulted in the later development of technology in the seven knowledge areas listed in Table 3. In other words, in respect to these knowledge areas, it took considerable time for technology to catch up with the advances in science. Our observations suggest that, for at least some important areas in DNA-NT, this field was science-intensive and challenging in the beginning. The possible reason for the ‘delay’ in technology might be the lack of complementary technologies at that time. Inventors would still have to do many experiments, learn from theories, and try out attempts in practice, to be skilful enough to realise their ideas into physical nanostructures and reach the required level of stability.

Table 4 shows the list of **technology-pull** terms we found in our data. There are 11 of such terms, a considerably larger number than the 7 science push terms we found earlier. Among the 11 terms found, ‘hybridisation chain reaction’ is the one with the longest time gap between technology and science. Not only being driven by techniques (DNA hybridisation, DNA detection, magnetic resonance imaging), and applications (biosensor, drug delivery), technology also took the lead in ‘programmability’ and ‘functionalization’, which turns structures into devices. To build machines at the nanoscale, technology signalled what it needs from science: ‘mechanical properties’.

Table 4. List of terms with strong signals for technology-pull

	Term	Threshold reached of 5% total frequency of each term in publications (1)	Threshold reached of 5% total frequency of each term in patents (2)	Time gap between (1) and (2)
1	Hybridization chain reaction	2010	1998	12
2	Functionalization	2001	1994	7
3	Liquid phase	1997	1990	7
4	Programmability	2001	1994	7
5	Biosensor	1999	1993	6
6	DNA detection	2001	1996	5
7	DNA hybridization	1999	1994	5
8	Drug delivery	2001	1996	5
9	Magnetic resonance imaging	1998	1993	5
10	Mechanical properties	1999	1994	5
11	Nucleic acid amplification	2001	1996	5

The areas marked with ‘technology-pull’ mostly occurred later than the effect of ‘science-push’[8]. The technology-pull areas are also the areas where science lagged behind technology. Some of ‘technology-pull’ knowledge areas are closely related to medical healthcare technologies, such as biosensor and magnetic resonance imaging. The long investment process of firms and other actors in those areas have resulted in patented inventions, which are not easily disclosed and diffused back to the general public, and academia. These areas may be the areas that scientists needed time to work, to collaborate with industry and applying them to their research publications. Especially, when ‘science of the artificial’ is concerned, technology comes first in the form of workable structures, devices and artefacts, which later become the object of scientific research. Furthermore, it is also worth to notice that the laboratory works always involve some equipment, which was patented several years ago. Traditional enabling techniques, methods used in the long-existing knowledge fields (such as molecular biology and biotechnology) are still usable/recombined in emerging DNA Nanoscience.

While the 7 science-push and 11 technology-pull terms discussed above are interesting, it is important to note that even a much higher number of terms, 73, does not show much time difference between the science and the technology domain. These represent the core areas where knowledge emerged and developed almost simultaneously, and are strong candidates for a case of *S&T co-evolution*. Scientific and technological knowledge might originate from the same place, the same person, or result from the co-creation process of scientists and inventors. Box 1 shows examples of some of these 73 areas. Among several enabling tools (microscopes) that we tested via occurring terms, both S&T have chosen atomic force microscope as the most-used tool during those 25 years. Other microscopes such as cryo-electron microscope, scanning tunnelling microscope, transmission electron microscope do not belong either to this category of co-evolution, or the views of ‘science-push’ and ‘technology-pull’. We think it might be the result of the selection mechanism going on between S&T.

Box 1. Example of terms reaching 5% threshold at approximately the same time

Cancer diagnosis, cancer treatment, logic gate and circuit, three-dimensional structure, nucleic acid structure, protein structure, living cell, DNA synthesis, DNA origami, DNA amplification, atomic force microscope, gold nanoparticle, self-assembly, liquid crystal

5.2 The evolution of knowledge content

Before discussing our quantitative results, we provide some examples of similar complementary and different terms, as we find them in Table 5. Examples of similar terms are ‘liquid crystal’, ‘mass spectrometry’ and ‘carbon nanotube’; They appear in the full period in both corpora. Examples of complementary terms are ‘cancer diagnosis’ paired with ‘cancer cell’, as well as ‘therapeutic agent’ paired with ‘drug delivery’. This table also shows examples of differences. For instance, in Sub-period 1, the term ‘microfluidic device’ only appears in patents, but the term ‘crystal structure’ only appears in publications.

[8] Because the moment that a term reached the 5% threshold Table 6 is often later than the moment that a term reached the 5% threshold in Table 5.

Table 5. Examples of similarity, complementarity[9], and difference in frequencies

		Full period	Sub-period 1 (1990-1997)	Sub-period 2 (1998-2005)	Sub-period 3 (2006-2015)
1	**Similar**	Liquid crystal, mass spectrometry, living cell, carbon nanotube, biological material	Self-assembly, mass spectrometry, molecular structure, binding affinity, polymer synthesis	Mass spectrometry, polymerase chain reaction, synthetic DNA, polymer synthesis, physical properties, nucleic acid structure	DNA origami, DNA synthesis, carbon nanotube, drug delivery, hairpin structure, , programmability, living cell
2	**Directly complementary**	[therapeutic agent, drug delivery], [therapeutic agent, cancer treatment], [x-ray chrystallography, protein structure], [programmability, design]	[functionalization, drug delivery], [design, biological molecule], [dna sequencing, dna fragment], [oligonucleotide, design]	[living cell, invivo], [structural analysis, molecular structure], [magnetic resonance imaging, magnetic properties]	[DNA origami, atomic force microscope], [gold nanoparticle, DNA origami], [cancer diagnosis, cancer cell], [self-assembly, functionalization]
4	**Different** (only in publications)	RNA structure, natural DNA, g-quadruplex DNA, two-dimensional structure	cryoelectron microscope, transmission electron microscope, crystal structure, carbon nanotube, structural stability, gold nanoparticle	cryoelectron microscope, molecular machine, molecular dynamic simulation, RNA detection, g-quadruplex DNA,	Scanning tunneling microscope, nuclear magnetic resonance, natural nucleic acid, RNA structure, g-quadruplex DNA
5	**Different** (only in patents)		microfluidic device	hybridization chain reaction	nucleic acid array
6	**Absent in both domains**		DNA origami, hybridization change reaction,short hairpin rna, bio stability, RNA interference, semiconductor device		

Table 6 - Row 1, provides the results of our quantitative analysis of commonality. The results show that the overall commonality level is high: 102 out of 110 terms (92%) are common (i.e. appearing in both domains) during the full period. Looking at the sub-periods, we observe that the commonality level increases over time, from 65% in Sub-period 1 to 89% in Sub-period 3. Row 2 shows our results from our similarity test for common terms that have the same relative frequency, using the chi-square test. Following Rayson & Garside (2000), we removed the terms with frequencies below 5, leaving us 91 terms to test. Our results reveal that only 13 terms are present in both domains with similar frequencies, while 78 terms are occurring in both domains but with different frequencies. Thus, out of all common terms that have the actual frequencies more than 5 occurrences, 14.3% also have similar relative frequencies. When considering the different sub-periods, that percentage is a bit higher, around 20%. Interestingly, our list of similar terms mostly overlaps with the lists of terms – candidates for co-evolution resulted from our analysis in Section 5.1. Those similar terms seemed to be mostly established knowledge areas in both domains, such as liquid crystal, biological material.

The increasing and high commonality, stable and low similarity, and increasing and high complementarity between two domains might suggest that scientific and technological knowledge in DNA-NT share the majority of the components, or building blocks, or knowledge areas. However, the knowledge structure and knowledge development mechanisms in both domains might be distinctive, but largely complementary in all sub-periods. This may result from

[9] The results of complementarity are to be updated.

the differences in nature of scientific and technological languages (terminologies), or various knowledge recombination processes going on in each domain.

In terms of differences, Table 6 - Row 3 and 4 show that that degree is low. Over the full period, there are only 8 terms that appear in publications but not at all in patents, and 0 terms the other way around. Considering different sub-periods, that number is a bit higher, but decreasing over time, suggesting convergence.

Table 6. A summary of common, similar and different terms[10]

		Full period	Sub-period 1 (1990-1997)	Sub-period 2 (1998-2005)	Sub-period 3 (2006-2015)
1	Common terms	102/110 (92%)	72/110 (65%)	97/110 (88%)	98/110 (89%)
2	Similar terms (excluding terms with less than 5 occurrences; Chi-square test: significant at p-value <0.05)	13/91 (14.3%)	9/40 (22.5%)	14/73 (19.2%)	16/67 (23.9%)
3	Different terms, only in publications	8/110	26/110	12/110	12/110
4	Different terms, only in patents	0/110	1/110	1/110	1/110
5	Absent in both domains	0/110	6/110	0/110	0/110

In all periods, we found evidence of the potential cognitive complementarity between two domains via 133 pairs of directly complementary terms. The number of indirectly complementary terms are much larger (10.525) than the number of directly complementary terms. Because all the 133 identified directly complementary terms are indirectly complementary, we decided to first analysis the dynamics of these 133 terms over time. Considering those 133 terms as 100% in the whole period, we found that the level of complementarity increased over time from 64% in Period 1 to 95% in Period 2 and 100% in Period 3.

Referring to Table 5, we realize that some terms (e.g. DNA origami) can be similar and complementary. If that happened, the term enjoyed its extensive growth in both domains. If only complementarity happened, often when the term does not belong to the mainstream of DNA-NT, it still embraces full potential of growth and dynamics in the next period.

6. Conclusions, research contributions, and research implications

Taking the case of DNA Nanoscience/Nanotechnology as a representative case for a promising and emerging knowledge area, this study finds that:

- The stylistic ‘science push’ and ‘technology pull’ patterns are not very dominant in our data. Only 7 out of the 110 specific knowledge subfields we studied follows a science-push pattern, and only 11 of 110 subfields follows a technology-pull pattern. So, even though relatively rare, technology-pull is stronger than science-push (in contrast to what has been assumed about nanotechnology in earlier literature, e.g. Zhao & Guan, 2013; and the smaller size of the technological domain of DNA-NT).
- The majority of the subfields in this area follows a pattern that is consistent with a *co-evolution* view of the relation between S&T.

[10] The results of complementarity are to be updated.

- In terms of knowledge content, there is a remarkably high degree of commonality (92%) between S&T. Although the level of knowledge similarity is quite low (14.3%), and remains stable, the level of complementarity increased over the three sub-periods.
- Similarity and complementarity are not mutually excluding concepts. Knowledge areas (such as DNA origami) which are both similar in both domains and complementary with other areas, have higher momentum to develop.

Furthermore, this study has several specific contributions in terms of methodology. Firstly, it shows how the ‘concept approach’ can be used to create datasets of publications and patents that have a high degree of both recall and precision, even in emerging technology areas where more standard approaches with keyword selection and classification searches do not work very well. Secondly, this study introduced a method of measuring (cognitive) complementarity, a topic that was until now neglected in the empirical literature on the link between S&T.

Finally, our study has policy implications: Given the small role of science-push patterns we observe in the field of DNA Nanotechnology, our study suggests that policies or interventions aimed purely at developing scientific knowledge are not likely to have much impact. The same thing holds for policies or interventions aimed purely at technology development. Policies that recognize that science and technology mostly developed hand in hand, following related, similar patterns, have a higher likelihood to have an impact.

7. Research limitations and future research

One important question is the degree to which our results can be generalized. We believe that the field of DNA Nanotechnology we studied has features that are typical for emerging, fast-growing and science-intensive knowledge fields, and that our results indeed can be generalised to such fields. Future studies, however, could test whether this belief holds.

Furthermore, our work has several more specific limitations. *First*, we assumed that knowledge in patents is equivalent to technological knowledge, knowledge in publications is equivalent to scientific knowledge. This is sometimes not true because not all scientific knowledge is represented in publications, or in the publication database we used (Web of Science), and, arguably more important, not all technological knowledge is embodied in patents (think of trade secrets). We do not have the data on funding and investment in the field, therefore, cannot offer any policy evaluation or suggestion on which domain develops more effectively based on the current funding regime.

Second, our data and analysis for the three research questions were text-based. We chose to ignore institutional factors affecting the development of DNA-NT and the citation relations between patents and publications. However, our study truly reflects the theories and practices described in the patents and publications in DNA-NT.

Third, while our approach allows us to identify patterns that are consistent with the view of co-evolution, it does not yet provide direct evidence for such co-evolution. That would require more direct observations of interaction between the two domains, and how one domain forms a selection mechanism for the other, and vice versa. One of the main challenges is that data on such interactions are scarce and notably incomplete. While Non-Patent Literature (NPL) references in patent data offer some data of the flow of knowledge from the science domain to the technology domain, no good indicator captures knowledge flows in the opposite direction.

The last limitation we want to mention here concerns the precision of our datasets (i.e. the proportion of selected records that are relevant). While we performed many incremental improvements to our query to improve both recall and precision, we encountered diminishing improvements and eventually decided to give a higher priority to recall than to precision (see above). In the future, when this knowledge area is more mature and boundaries became more evident, this precision may be increased.

Regarding *future research*, we plan to address the first limitation mentioned above by new work that looks at tacit knowledge in the context of co-evolution of knowledge. Another work could compare the knowledge recombination in each domain. We also plan future research that further investigates the speed of development of specific subareas in the field of DNA Nanoscience/Nanotechnology, with the aim to help in formulating science, technology and innovation policies.

Annex A: The four concepts used in the concept approach

Concept A: Nanotechnology	'Nanotechnology is science, engineering, and technology conducted at the nanoscale, which is about 1 to 100 nanometers' (Definition from the US National Nanotechnology Initiative's website, see Reference). Therefore, any science or technology that works below the scale of 100 nanometers is considered 'nanotechnology'. This definition is a broad one. Therefore we could maximise the search in this sub-concept by collecting synonyms referring to nanoscale or instruments that have been used in nanotechnologies, such as specific types of microscopes (AFM, TEM, SEM…).
Concept B: Design	The word 'design' has two forms, the verb and the noun. According to Miller (2005), as a noun, 'design' refers to an object or an entity. As a verb, it refers to a process or series of activities. Design is the construction of an object or creation of an entity. An interesting feature of the DNA origami technique is that DNA strands are programmed, synthesized, and can self-assemble themselves afterwards. I found all terms related to this process and listed them in the sub-concept 'design'.
Concept C: Structure	Structure is defined by the Oxford Dictionary as 'a particular arrangement of parts'. In this sub-concept, I found several synonyms of 'structure' based on publications and patents in DNA-NT by top contributors in the field such as Nadrian Seeman, Paul Rothemund, and others. We also noted specific words that are related to DNA structures and included in the search.
Concept D: DNA	DNA is the abbreviation of 'deoxyribonucleic acid', a type of nucleic acid, a chemical that carries genetic information in the cells of animals and plants (Oxford Dictionary), or any living organisms, and viruses (Wikipedia). It is interesting to note that the term DNA, used in our research refers to artificial DNA, not its natural form. However, its synonyms and related terms are borrowed from molecular biology.
Exclusion terms in Titles (E1)	At the first level of exclusion, we excluded specific terms related to other closely related fields (e.g. molecular biology, genetic engineering, forensics…). However, these terms still have chances to appear in abstracts or keywords.
Exclusion terms in Titles, Abstracts and Keywords (E2)	At the second level of exclusion, we excluded the terms that should never appear in titles, abstracts and keywords. This is the strongest exclusion to improve the precision of our data.

Final query: (**Nanotechnology** AND **Design** AND **Structure** AND **DNA**) NOT (E1 OR E2)

.. where

Nanotechnology = NANO* OR 'ATOM* FORCE MICROSCOP*' OR AFM OR TEM OR 'TRANSMISSION ELECTRON MICROSCOP*' OR SEM OR 'SCANNING ELECTRON MICROSCOP*' OR 'FLUORESCENCE MICROSCOP*' OR 'CRYO-ELECTRON MICROSCOP*' OR 'CRYO-EM' OR MOLECUL* OR MULTIMER$ OR MONOMER$

Design = DESIGN* OR COMPUT* OR CONJUGAT* OR FORM* OR FOLD* OR JUXTAPOS* OR PROGRAM* OR BIND* OR BOUND OR ATTACH* OR LINK* OR CONNECT* OR CONSTRUCT* OR BRANCH* OR BOND* OR FABRICAT* OR 'SELF-ASSEMBL*' OR 'SELF-REPLICAT*' OR 'SELF-ORGANI*' OR 'DIRECTED-ASSEMBL*' OR SYNTHETIC OR ARTIFICIAL OR 'NON-NATURAL' OR UNNATURAL OR 'NON-GENETIC'

Structure ='*STRUCTURE$' OR DOMAIN$ OR SYSTEM* OR MOTOR* OR MACHIN* OR DEVICE$ OR ARRAY$ OR POLYHEDR* OR CONJUGATE$ OR LADDER$ OR '*ROBOT*' OR JUNCTION$ OR SCAFFOLD* OR TEMPLAT* OR TILE$ OR TILING$ OR LATTICE$ OR 'STICKY END*' OR 'COHESIVE END*' OR STAPL* OR 'LOGIC GATE*' OR CIRCUIT$ OR ORIGAMI

DNA = DNA* OR '*NUCLEIC ACID*' OR 'DOUBLE HELI*' OR HELICES OR '*STRAND*' OR '*NUCLEOTID*' OR FOLDAMER$ OR APTAMER$

E1 = CELL$ OR THERAP* OR INFLAMMAT* OR RIBOSOME$ OR BODY* OR SPECIES* OR BRAIN* OR 'MOLECULAR CLONING' OR EVOLUTION* OR IMMUN* OR DISORDER$ OR VIRUS* OR ORGANISM$ OR ORGAN$ OR BACTERI* OR ANTIBOD* OR HUMAN* OR MAMMAL$ OR TISSUE$ OR TRANSCRIPTION OR RAT$ OR MICE OR HSP* OR P53 OR STAT3 OR 'RIBONUCLEIC' OR 'NON-NUCLEIC' OR 'DNA SEQUENCING' OR 'GENETIC ENGINEERING' OR GENETICS OR SYMPTOM$

E2 = METABOL* OR GEOGRAPH* OR NUTRI* OR YEAST$ OR TREE$ OR SOIL OR FISH* OR MARINE OR INJUR* OR WOUND* OR 'GENE EXPRESSION' OR 'GENETIC STRUCTURE' OR 'GENETICALLY MODIFIED' OR GMO OR 'GENETICALLY ENGINEERED' OR 'GENE REGULATION$' OR 'GENETIC ALGORITHM$' OR 'GENE DELIVERY' OR 'GENE INTERACTION$' OR 'GENO*' OR 'PHYLOGEN*' OR TRANSGENIC OR HORMON* OR ESTROGEN OR TESTOSTERONE OR PATIENT$ OR EMBRYO* OR POLYMERASE OR VACCIN* OR ANTIBIOTIC$ OR BLOOD OR FETAL OR FETUS OR OFFSPRING$ OR BLAST OR FUNG* OR MUTAT* OR CHROMOSOME OR 'PRO POLYPEPTIDE$' OR HELICASE OR INFECT* OR INSECT* OR PLANT$ OR ANIMAL$ OR FORENSIC$ OR NANOPLANKTON OR NANOFAUNA OR CAS9* OR NANO2 OR NANO3

References

Arora, S.K., Porter, A.L., Youtie, J., & Shapira, P. (2013). Capturing new developments in an emerging technology: An updated search strategy for identifying nanotechnology research outputs. *Scientometrics*, *95*(1), 351–370.

Arthur, W.B. (2009). *The nature of technology: What it is and how it evolves*. New York: Free Press.

Audretsch, D.B., Bozeman, B., Combs, K.L., Feldman, M., Link, A.N., Siegel, D.S., Stephan, P., Tassey, G. (2002). The Economics of Science and Technology. *Journal of Technology Transfer*, *27*(2), 155–203.

Bekkers, R., & Martinelli, A. (2012). Technological Forecasting & Social Change Knowledge positions in high-tech markets : Trajectories , standards , strategies and true innovators. *Technological Forecasting & Social Change*, *79*(7), 1192–1216.

Benson, C.L., & Magee, C.L. (2013). A hybrid keyword and patent class methodology for selecting relevant sets of patents for a technological field. *Scientometrics, 96*(1), 69–82.

Benson, C. L., & Magee, C. L. (2015). Technology structural implications from the extension of a patent search method. *Scientometrics*, *102*(3), 1965–1985.

Bernardes, T.A.; Albuquerque, E.M. (2003). Cross-over, thresholds, and interactions between science and technology: lessons for less-developed countries. *Research Policy*, *32*(5), 865–885.

Blondel, V. D., Guillaume, J., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment 2008* (10), P1000.

Boon, M. (2009). Instruments in science and technology. In: Jan-Kyrre Berg Olsen; Stig Andur Pedersen; Vincent F. Hendricks (Ed.), A Companion to Philosophy of Technology. Wiley-Blackwell, pp. 78–84.

Boon, M. (2009). Instruments in science and technology. In Jan-Kyrre Berg Olsen; Stig Andur Pedersen; Vincent F. Hendricks (Ed.), *A Companion to Philosophy of Technology* (Blackwell, pp. 78–84). Wiley-Blackwell.

Boschma, R., & Frenken, K. (2009). *Technological relatedness and regional branching*. Retrieved from http://econ.geo.uu.nl/peeg/peeg.html

Boschma, R., Heimeriks, G., & Balland, P. (2014). Scientific knowledge dynamics and relatedness in biotech cities. *Research Policy, 43*(1), 107–114.

Boyack, K. W. ., Klavans, R. ., & Börner, K. (2005). Mapping the backbone of science. *Scientometrics*, *64*(3), 351–374.

Breschi, S., & Catalini, C. (2010). Tracing the links between science and technology: An exploratory analysis of scientists' and inventors' networks. *Research Policy*, *39*(1), 14–26.

Bush, V. (1945). *Science, the endless frontier: report to the President by Vannevar Bush, Director of the Office of Scientific Research and Development, July 1945*. United States Government Printing Office, Washington.

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. *Scientornetrics*, *22*(1), 155–205.

Choi, S., Yoon, J., Kim, K., Yeol, J., & Kim, L. C. (2011). SAO network analysis of patents for technology trends identification : a case study of polymer electrolyte membrane technology in proton exchange membrane fuel cells, *Scientometrics*, *88*, 863–883.

Darby, M. R., & Zucker, L. G. (2003). *Grilichesian breakthroughs: inventions of methods of inventing and firm entry in nanotechnology*. NBER working paper No. 9825.

Dosi, G. (1982). Technological paradigms and technological trajectories. *Research Policy*, *22*(2), 102–103.

Dosi, G., Llerena, P., & Sylos, M. (2006). The relationships between science , technologies and their industrial exploitation : An illustration through the myths and realities of the so-called ' European Paradox '. *Research Policy, 35*(10), 1450–1464.

Douglas, K. (2016). *DNA Nanoscience: From Prebiotic Origins to Emerging Nanotechnology*. CRC Press.

Edquist, C., & Hommen, L. (1999). Systems of innovation: Theory and policy for the demand side. *Technology in Society, 21*(1), 63–79.

Engelsman, E. C., & van Raan, A. F. J. (1992). A patent-based cartography of technology. *Research Policy*, *23*(1), 1–26.

European Commission (1995). Green Paper on Innovation. COM (95) 688 final, 20 December 1995. Bulletin of the European Union Supplement, 5/95. European Commission: Brussels.

Evert, S. (2005). *The Statistics of Word Cooccurrences Word Pairs and Collocations*. Institut für maschinelle Sprachverarbeitung, Universität Stuttgart. Retrieved from http://elib.uni-stuttgart.de/bitstream/11682/2573/1/Evert2005phd.pdf

Gil-Pérez, D., Vilches, A., Fernández, I., Cachapuz, A., Praia, J., Valdés, P., Salinas, J. (2005). Technology as ' Applied Science '. *Science & Education, 14*(3–5), 309–320.

Han, F.; Magee, C. L. (2017). Testing the science/technology relationship by analysis of patent citations of scientific papers after decomposition of both science and technology. (April, 2017). Retrieved from https://arxiv.org/abs/1705.00258

Huang, Y., Schuehle, J., Porter, A. L., & Youtie, J. (2015). A systematic method to create search strategies for emerging technologies based on the Web of Science: illustrated for 'Big Data'. *Scientometrics,* 105(3), 2005–2022.

Joung, J., & Kim, K. (2017). Technological Forecasting & Social Change Monitoring emerging technologies for technology planning using technical keyword based analysis from patent data. *Technological Forecasting & Social Change*, *114,* 281–292.

Kallis, G. (2007). Socio-environmental co-evolution: some ideas for an analytical approach. *International Journal of Sustainable Development*, *14*(1), 4–13.

Kallis, G., & Norgaard, R. B. . (2010). Coevolutionary ecological economics. *Ecological Economics,* 69, 690–699.

Kilgarriff, A. (2001). Comparing corpora. *International Journal of Corpus Linguistics, 6*(1), 97–133.

Klevorick, A. K., Levin, R. C., Nelson, R., & Winter, S. G. (1995). On the sources and significance of interindustry differences in technological opportunities. *Research Policy, 24*(2), 185–205.

Kuhn, T. S. (1970). *The structure of scientific revolutions* (2nd ed.). University of Chicago Press.

Layton, E. T. (1974). Technology as Knowledge. *Technology and Culture, 15*(1), 31–41.

Leydesdorff, L. (1994). The evolution of communication systems. *International Journal of Systems Research and Information Science,* 6 (1994), 219-230

Makri, M.; Hitt, M.A.; Lane, P.J. (2010). Complementary technologies, knowledge relatedness, and invention outcomes in high technology mergers and acquisitions. *Strategic Management Journal*, *31*(6), 602–628.

Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy*, *29*(3), 409–434.

Mina, a., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy*, *36*(5), 789–806.

Mowery, D., & Rosenberg, N. (1979). The influence of market demand upon innovation: A critical review of some recent empirical studies. *Research Policy, 8*(2), 102–153.

Murray, F. (2002). Innovation as Co-evolution of Scientific and Technological Networks: Exploring Tissue Engineering. *Research Policy, 31*(2002), 1389–1403.

Nakagawa, H. (2000). Automatic Term Recognition based on Statistics of Compound Nouns. Terminology. *International Journal of Theoretical and Applied Issues in Specialized Communication*, *6*(2) 195–210.

Narin, F., Hamilton, K.S., & Olivastro, D. (1997). The increasing linkage between U.S. technology and public science. *Research Policy, 26*(3), 317–330.

Narin, F., & Noma, E. (1985). Is technology becoming science? *Scientometrics*, *7*(3), 369–381.

Nelson, R. R.; Rosenberg, N. (1993). Technical innovation and national systems. In. Nelson R.R (Eds.), *National innovation systems: A comparative analysis*. New York: Oxford University Press, pp. 3–21.

Nordmann, A. (2008). Philosophy of NanoTechnoScience. In: G. Schmid, H. Krug, R. Waser, V. Vogel, H. Fuchs, M. Grätzel, K. Kalyanasundaram, L. Chi (Eds.) *Nanotechnology,* Vol. 1: G. Schmid (Eds.) Principles and Fundamentals. Weinheim: Wiley, 2008, pp. 217-244.

Patra, D. (2011). Nanotechnoscience, nanotechnology, or nanotechnoscience: Perceptions of Indian nanoresearchers. *Public Understanding of Science, 22*(5), 590–605.

Pavitt, K. (1987). The objectives of technology policy. *Science and Public Policy*, *14*, 182–188.

Perkmann, M. (2009). The two faces of collaboration: impacts of university- industry relations on public research. *Industrial and Corporate Change*, *18*(6), 1033–1065.

Pinheiro, A.V., Han, D., Shih, W.M., & Yan, H. (2011). Challenges and opportunities for structural DNA nanotechnology. *Nature Nanotechnology, 6*(12), 763–772.

Porter, A.L., Youtie, J., Shapira, P., & Schoeneck, D.J. (2008). Refining search terms for nanotechnology. *Journal of Nanoparticle Research*, *10*(5), 715–728.

de Solla Price, D. J. (1965). Is technology historically independent of science? A study in statistical historiography. *Technology and Culture, 6*(4), 553–568.

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In: *Proceedings of the workshop on Comparing corpora - Volume 9*. Association for Computational Linguistics, pp. 1–6.

Rosenberg, N. (1982). How exogenous is science. In: N. Rosenberg, *Inside the Black Box: Technology and Economics* (pp. 141–160). Cambridge University Press.

Rothemund (2006). Folding DNA to create nanoscale shapes and patterns. *Nature*, *440*(7082), 297–302.

Schmoch, U. (2007). Double-boom cycles and the comeback of science-push and market-pull. *Research Policy, 36*(7), 1000–1015.

Sedighi, M. (2016). Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of Informetrics). *Library Review*, *65*(1/2), 52–65.

Simon, H.A. (1969). The sciences of the artificial. Cambridge: MIT Press.

Xu, G., Zhang, Y., & Li, L. (2011). *Web Mining and Social Networking*. Springer Science+Business Media.

Stefano, G. D., Gambardella, A., & Verona, G. (2012). Technology push and demand pull perspectives in innovation studies : Current findings and future research directions. *Research Policy, 41*(8), 1283–1295.

Tahmooresnejad, L., & Beaudry, C. (2015). Does government funding have the same impact on academic publications and patents? The case of nanotechnology in Canada. *International Journal of Innovation Management, 19*(3), 1–16.

Van Den Bergh, J.C.J.M., & Gowdy, J.M. (2000). Evolutionary Theories in Environmental and Resource Economics Approaches and Applications. *Environmental and Resource Economics, 17*(1), 37–57.

Van Den Bergh, J.C.J.M., & Stagl, S. (2003). Coevolution of economic behaviour and institutions : towards a theory of institutional change. *Journal of Evolutionary Economics 13*(3), 289–317.

Yu, C., & Yingming, L. (2008). *Co-evolution of Technological Regime and Industrial Clusters: The Catching-up of Dounan's Flower Industry in China*. Paper presented in the IV Globelics Conference at Mexico City, September 22-24, 2008.

Zhang, J., Xie, J., Hou, W., Tu, X., Xu, J., Song, F., … Lu, Z. (2012). Mapping the Knowledge Structure of Research on Patient Adherence : Knowledge Domain Visualization Based Co-Word Analysis and Social Network Analysis, PLoS ONE, *7*(4), 1–7.

Zhao, Q., & Guan, J. (2013). Love dynamics between science and technology: some evidences in nanoscience and nanotechnology. *Scientometrics*, *94*(1), 113–132.

Zucker, L. G., & Darby, M. R. (2005). *Socio-economic impact of nanoscale science: Initial results and nanobank*. NBER working paper No.11181.

[a] It is worthwhile noticing that some records, in which the sub-concept 'nanotechnology' is implicit, should be included in our datasets. Certain authors or inventors chose not to mention nano-related terms explicitly, rather than discussing only DNA or oligonucleotides. That might be the reason why a considerable number of patents belonging to DNA-NT is not classified under IPC-code B82 (Nanotechnology). From a conceptual point of view, DNA and nano are quite different. However, from a practical point of view, when one discusses DNA or nucleotides, one should infer that the research is conducted at the nanoscale, since the dimension of a DNA strand is approximately 2.5 nanometers. Therefore, 'DNA' and '*nucleotid*' are included in two concept areas to avoid missing certain records that do not mention nano-related terms. Although DNA and Nanotechnology are closely related concepts, they were not grouped because it led to more considerable noise in the datasets selected. Thus, DNA-related terms must appear in the set under any conditions, while the presence of nano-related terms remains an option.