Science and Technology Relatedness: The Case of DNA Nanoscience and DNA Nanotechnology

Hanh Luong La and Rudi Bekkers

Abstract The relatedness between knowledge components within the science domain is widely discussed in the economic, innovation, and management literature. The same is true for the technology domain. Yet, the relatedness between knowledge components across these knowledge domains has received considerably less attention. This chapter aims to introduce the concept of knowledge relatedness between science and technology (S&T), which have been disentangled as two distinct corpora. We approach S&T relatedness from two perspectives: content relatedness (with four indicators: similarity, complementarity, commonality, difference) and temporal relatedness. We then test our ideas with novel empirical material from the field of DNA nanoscience and DNA nanotechnology. We find that the relatedness between S&T scores relatively low, which may explain the relative lack of commercial activity in this field. In light of their indirect complementarity, we recommend that funding "bridging areas" could lead to simultaneous progress in S&T.

Keywords Science and technology relatedness · Knowledge relatedness · Knowledge complementarity · Concept approach · Text-mining · DNA nanotechnology

JEL Code O33

H. L. La (✉)
UNU-MERIT, Maastricht University, Maastricht, The Netherlands

Faculty of Economics, Academy of Finance, Ha Noi, Viet Nam

School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands
e-mail: la@merit.unu.edu; krasplanet@gmail.com

R. Bekkers
School of Innovation Sciences, Eindhoven University of Technology, Eindhoven, The Netherlands

A. Pyka, K. Lee (eds.), *Innovation, Catch-up and Sustainable Development*, Economic Complexity and Evolution, https://doi.org/10.1007/978-3-030-84931-3_3

1 Introduction

The relation between science and technology (S&T), two knowledge domains that are believed to be main sources for innovation and economic growth,[1] is a broad and fascinating topic for evolutionary economists and science-technology-innovation (STI) policymakers. It is widely accepted that S&T are interacting, interdependent, and interconnected entities (Breschi & Catalini, 2010; Meyer, 2000; Wang & Li, 2018), especially in science-based technologies. The nature of S&T relationship, therefore, can be investigated in a narrower sense via S&T interaction. Such interaction, for instance, between the public science sector and the private sector, is a crucial factor shaping the competitiveness of firms, regions, and countries (Nomaler & Verspagen, 2008). However, S&T interaction cannot easily be observed directly. Most empirical literature studies S&T interaction by looking at similarities (e.g., patent-paper pairs[2]) and at linkages (scientific non-patent literature[3]). Possible complementarities between these domains have received relatively little attention.

Addressing the gap in both theory and empirics, this chapter introduces the concept of "S&T relatedness" as a proxy for S&T interaction. It is an umbrella concept encompassing both similarities and complementarities across the domains. We theorize that the higher the S&T relatedness (but not only S&T similarities), the more economically one can further develop both domains, given the scarcity of resources, including funding R&D projects. A higher S&T relatedness also means a higher probability that a scientist in the field reaches out of her specialization towards technology-oriented activities, or a higher probability that an inventor in the field engages in more science-based activities. Via measuring S&T relatedness empirically, we aim to find which knowledge areas in both domains should deserve more attention. Choosing a text-mining and keyword analysis approach, we aim to identify the most important knowledge areas and their relatedness across S&T domains.

We tested our concept on the case of DNA nanoscience and DNA nanotechnology (to which we will from now on refer to as DNA-Nano). We found this field is growing in science, and promises many emerging technological applications (e.g., in electronics, molecular and cellular biophysics, biomimetic systems, energy transfer and photonics, and in diagnostics and therapeutics for human health, Pinheiro et al., 2011). However, actual industrial applications are lagging behind, and there has been little marketable activity (Dunn, 2020). We suspect if there was due to too little S&T interaction, or a significant technology lag in comparison to science. We asked ourselves: "How closely related is the knowledge in both S&T regarding this specific field?", "How can one enhance the growth of both S&T economically?", and

[1] See the discussion on neo-classical and evolutionary theories in Nelson and Winter (1974) and concerns raised by Dosi (1982), Suenaga (2015), and others about uncertainties related to S&T that may cause new technological paradigms.

[2] We later refer to these as PPPs.

[3] We later refer to it as NPL.

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

	Literature that considers S&T as two distinct entities	Literature that focuses on S&T convergence
Theoretical	I	II
	Dosi (1982), Pavitt (1987), Price (1965)	Arthur (2009), Layton (1974), Nordmann (2008)
Empirical	IV	III
	Mina et al. (2007), Zhao and Guan (2013)	Breschi and Catalini (2010), Murray (2002)

"Which knowledge areas in S&T should deserve more priorities for funding and development"?

The remainder of the chapter is structured as follows: we start by discussing the current literature on the S&T relationship, S&T interactions, and on knowledge relatedness in Sect. 2, as well as introducing our research questions. Then, Sect. 3 presents our research methods, including the selection of S&T domains and the measurement of S&T relatedness. We measure S&T relatedness in two dimensions: content relatedness and temporal relatedness. In Sect. 3, we also present an overview of our empirical data. Section 4 discusses our results, while Sect. 5 offers discussion and conclusion.

2 Literature Review: From the S&T Relationship to S&T Relatedness

2.1 S&T Relationship and Interaction

The S&T relationship and interaction is a recurring and fascinating topic in the economic and innovation literature. It can be considered an interrelationship, because multiple knowledge components in science are connected to multiple knowledge components in technology, and we encounter variations across these S&T domains. We discuss the theoretical and the empirical literature that focuses on observable patterns in the development of S&T. We conclude the section by raising our research questions.

Since S&T are very much interrelated, numerous works have focused on comparing their knowledge developments. While both domains encompass research activities, their objectives are different. Science aims to discover, describe phenomena, and build theories (Drexler, 2013, p 116; Kuhn, 1970, p 60). Technology aims to find solutions for problems and is more concerned with design and production (Dosi, 1982; Drexler, 2013, p 117).

Table 1 shows selected literature on comparing knowledge development between S&T. Basically, there are two main streams, and both acknowledge the interaction between the two domains. However, the first stream (Quadrant I and IV) considers

S&T as two separate entities, whereas the second stream (Quadrant II and III) regards them as two converging entities.

Quadrant I comprises the theoretical literature that considers S&T as two separate entities, typically characterized by similarities and complementarities. Examples of scholars who followed this approach are Price (1965), Dosi (1982), and Pavitt (1987). The work by Price (1965) is considered one of the earliest seminal studies on the S&T relationship and interaction. It refers to Toynbee's "pair of dancers" as a metaphor for the relationship between S&T. Price implies that S&T are two (parallel) co-evolving, cumulative, and autonomous structures/entities. Although the dancers could be men or women, with *differences* in attitude and structure, they move to the *same* music. In the view of Price, the "S&T dancers" typically have "infrequent interaction," a "separate cumulating structure" and more interestingly, are considered to be *complementary*. Two decades later, Dosi (1982) describes the two domains in terms of scientific and technological paradigms, and scientific and technological trajectories. He reiterates Thomas Kuhn's (1970) view of a scientific paradigm as a model, a pattern, and a set of problems of inquiry. In an analogy of Kuhn's scientific paradigm, Dosi defines the 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 argues that the efficiency of the whole field is not inevitably an outcome of creating more similarities between S&T. He emphasized that policymaking should consider the *complementarity* between S&T, which "varies considerably among sectors of application, in terms of the direct usefulness of academic research results, and the relative importance attached to such results and to training."

Quadrant IV comprises empirical studies that consider S&T as two separate entities, and is, compared to the other quadrants, understudied. Mina et al. (2007) study the evolution of scientific and technological knowledge on the treatment of coronary artery disease by comparing the two *top main paths*[4] of its scientific and technological citation networks and found them somewhat similar. From a different perspective, Zhao and Guan (2013) introduce a model characterizing the relationship between S&T based on their classification of S&T styles and the changes in producing publications and patents. While their approach is novel, their dataset (on nanotechnology) was limited to publications and patents at selected universities only. Their work thus ignores the role of industry in publishing and patenting.

Quadrant II comprises theoretical contributions investigating the S&T knowledge relationship via the integration or overlap between these domains. Layton (1974) explains how transforming a set of technological rules became a new entity

[4]The main path approach is a network analysis tool introduced in the late 1980s to investigate networks of scientific publications, and later to study patent networks (see Verspagen, 2007; Bekkers & Martinelli, 2012). The top main path is considered as representing the most important developments in citation networks.

of science: "technological science" or "engineering science." In a similar vein, Arthur (2009) further articulates that S&T are "deeply interwoven." In fact, in the field of nanoscience and nanotechnology, some scholars articulate the term "nanotechnoscience" (Nordmann, 2008; Patra, 2011). Such terms reflect the belief in a true integration of S&T, a context in which we cannot simply distinguish between S&T, or between basic and applied research (Nordmann, 2008).

Quadrant III comprises empirical contributions that examine the S&T knowledge relationship via the convergence or overlap between scientific and technological networks. Scholars in this quadrant emphasize similarities, rather than complementarities, making the differences between S&T appear insignificant (Meyer, 2000). Since Narin et al. (1997), a large body of quantitative literature used NPL references as a direct proxy for S&T interaction including Meyer (2000), Verbeek et al. (2002). Other studies, such as those of Murray (2002) and Chang et al. (2017), investigate S&T interaction via patent-paper pairs (PPPs), 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 can form the basis for further analysis. Murray's work (2002) forms the basis for Boyack and Klavans (2008), Breschi and Catalini (2010), who trace the link between scientific and technological networks via their gatekeepers: inventors-authors. Perhaps, the emerging topics around these gatekeepers are just the tip of the iceberg, reflecting only the part of both networks where the similarities are the strongest and most visible. Arguably, the S&T interaction may occur in certain other places than just where direct citation links, PPPs or inventors-authors exist, and the largest share of knowledge is through work by non-author inventors and non-inventor authors. If this is true, then it would be good to look at the S&T interaction also from a broader perspective, through various patterns of interaction (e.g., complementarities), rather than only based on similarities. We also note that observing citations links has inherent limitations: while patents do at some rate refer to scientific publications (NPL), scientific publications rarely refer to knowledge contained in patents, even if granted patents, by mere definition, must be novel.

2.2 From Knowledge Relatedness to S&T Relatedness

The literature on knowledge relatedness is fragmented and not well-established. The S&T relatedness and knowledge relatedness between two domains have not been discussed in any literature. In this sub-section we will discuss the "relatedness" as a "universal" concept, and then in different contexts, ranging from computational linguistics, management studies to economic geography, then explain why we need this concept in explaining S&T interaction.

Most of the literature refers to "relatedness" as the measure of proximity—or distance—between two entities, activities, or components, generally within one domain (in one corpus, in science or in technology, in one region, or in one sector, etc.). Originating from one domain, these entities normally are not identical but sharing some commonalities. The relatedness between two entities is often measured by the overlap (via co-classification or co-occurrences) between them. Therefore, knowledge relatedness has been mostly equated with knowledge similarity, which just reflects part of the whole picture of all possible patterns of relatedness. In computational linguistics, semantic relatedness is often used interchangeably with semantic similarity, which is the distance between two-word vectors (measured by the cosine of the angle between vectors, Euclidean distance, or Spearman rank correlation coefficient, etc.).

Economic geographers and innovation economists see technology relatedness as the extent to which the variety of technologies being used in a region is related (Boschma & Frenken, 2009). Scientific relatedness refers to the cognitive distance between a new potential scientific topic and a set of specialized topics (Boschma et al., 2014). These concepts of relatedness are often employed to study how specialization and diversity influence firms' performance or regional economic growth.

Makri et al. (2010) investigate science similarity and complementarity, technology similarity and complementarity, but only at a firm level. In this study, they conceptualized knowledge relatedness as knowledge similarity and complementarity. They argued that technological overlap can proxy the similarity of technological assets but cannot capture possible technological complementarities. Even 10 years after their publication, knowledge complementarity is still under-researched in different contexts.

As far as we are concerned, knowledge production is an interactive, path-dependent, and cumulative process (Boschma et al., 2014; Dosi, 1982). The extent to which knowledge entities are related can also reflect the interaction between agents. According to Tripodi et al. (2020), knowledge relatedness increases the probability of a scientist reaching out of her own specialization. Looking at our context of S&T relationship, S&T relatedness could indicate the probability of a scientist engaging in more technology-oriented activities or an inventor engaging in more science-based activities. It could also reflect the interactive learning process between scientists and inventors, in short S&T interaction.

In summary, the literature on the S&T relationship and interaction, and knowledge relatedness discusses both similarities and complementarities. The empirical literature, however, mostly focuses on similarities, sometimes on differences, and hardly focuses on complementarities. Empirical works on S&T similarities mainly use PPPs, as a proxy for S&T interaction. But we think there might be more room to discuss the S&T interaction in a more systematic manner, because PPPs just reflect the similarities in an incomplete extent.[5] The players in both S&T can interact (or learn from each other) in multiple ways[6] (for instance, reading and referring to others' work, but also being co-funded in the same project, or sharing the same

[5] In a similar vein, Heinisch et al. (2016) used co-location as a proxy for direct knowledge interaction.

[6] Both directly and indirectly.

equipment), which contribute to the similarities and complementarities. Moreover, empirical work on S&T relatedness, including S&T complementarity, remains a research gap in S&T studies and knowledge relatedness across domains. For these reasons, our study aims to introduce the concept “S&T relatedness,” its dimensions and measurement. In this chapter, we test it empirically on the case of DNA-Nano S&T. Thus, we aim to investigate empirically to what extent the knowledge contents in DNA Nanoscience and DNA Nanotechnology are related; more specifically, how they are similar, complementary, or different, over time. Additionally, we also look at the temporal relatedness of these domains, based on the gap between the emergence of knowledge areas in each domain.

3 Methods and Data

To study S&T relatedness, we consider these two domains as two corpora, i.e. bodies of text. In text-based methodologies, science is often proxied by academic publications,[7] whereas technology is often proxied by patents. By combining our relatedness metrics with text-mining publications and patents, we aim to discover narrative information within and across the two interrelated domains. Such a method is useful not only in information retrieval but also in the evaluation of research and funding, future complementary qualitative research, STI studies, and policymaking.

Accordingly, we extract publications (mainly journal articles) and patents systematically from two database platforms (Web of Science, provided by Clarivate Analytics and PATSTAT by the European Patent Office), which provide extensive search and retrieval facilities within their meta-data. Accordingly, we employ text-mining techniques to convert unstructured data (raw text) into structured data, namely “knowledge areas” represented by the most “significant” terms[8] (a smaller unit of analysis[9]).

In a nutshell, our methodology is four-fold: assembling two corpora, one for science and one for technology, by retrieving relevant documents from the respective databases, using our concept approach (Sect. 3.1), text-mining methods that extract key terms with their occurrences and co-occurrences from each corpus and can proxy the respective knowledge base underlying the two knowledge domains (Sect. 3.2), measuring the content relatedness between S&T by several indicators: commonality, similarity, complementarity (direct, indirect), and difference (Sect. 3.3), and measuring the temporal relatedness between S&T based on the emergence of knowledge areas (Sect. 3.4). In Sect. 3.5, we provide a description of our data.

[7]Note that while we use the term “academic publications,” such publications can also be authored by people working for firms. Likewise, university staff can also apply for patents.

[8]They are “term groups,” which consist of synonyms, abbreviations...which have the same meaning.

[9]We used two levels of analysis: domain level, and term level.

3.1 Selecting S&T Domains: The Concept Approach

For both publications and patents, the most common search/selection strategies are keyword search and classification search (Benson & Magee, 2013), or the combination of both. The keyword search typically employs 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 when patents are hierarchically classified according to technology/application areas (e.g., IPC or CPC codes). More sophisticated approaches for keyword search use 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 "artifact" component and a "knowledge" one (Benson & Magee, 2015).

The selection procedure to build the datasets of publications and patents is a critical step, and we evaluate our selection using two criteria: recall and precision. Recall is defined as the proportion of all relevant records retrieved, whereas precision is the proportion of retrieved records that are relevant. Both in practice and (information retrieval) theory, it is hard for any query to achieve perfect recall and precision at the same time, because of the inherent trade-off between the two. Search strategies can increase recall (e.g., using synonyms, wild-flags, and OR operators) at the expense of lower precision. Alternatively, search strategies can increase precision (e.g., using AND operators together with highly specific search terms) typically imply lower recall. The true challenge is to find an appropriate balance between recall and precision in a given context. The achievable levels of recall and precision also depend on the subject area and the novelty of the field. In emerging fields, tracking patents and publications is often challenging (Huang et al., 2015). Data might be poorly defined, and terminology may change over time. Classifications systems for publications/journals and for patents may not yet offer specific classes for emerging fields. The researchers often face the challenges of either low recall or low precision or the imbalance in the sub-areas of the emerging field (*ibid.*).

It is worth noting that for data retrieval in emerging fields, the requirement for precision is often considered to be not as important as in well-established fields. Porter et al. (2008) argue 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 is useful in emerging technology studies. We think Huang et al. (2015)'s suggestion above is quite reasonable and applicable in our case, because for an emerging field like DNA-Nano, it is harder to achieve precision than recall. While we can define and estimate recall by counting the presence of relevant contributions by key individuals in DNA-Nano, defining and estimating precision is a daunting and infeasible task. Among other things, this is because the boundaries of an emerging field with its adjacent fields have not yet been precisely defined.[10] Moreover, each individual expert in the field works within

[10] This may due to the fact there is no fixed perfect definition for a new field.

his/her narrow area of expertise and is not fully aware of the knowledge development and recombination in the entire field. The growth of the field now has much gone beyond what Seeman—the pioneer of DNA-Nano, and his first-generation students ever imagined. Based on the above considerations, for this study, we choose to prioritize recall over precision.

Our initial exercises with keyword search and classification search strategies for DNA-Nano (a field we will describe later) revealed low levels of recall and precision. Most likely, this was because it is an emerging, complex technology field, whose boundaries with other knowledge fields (e.g., bio-nanotechnology, biochemistry, biophysics) are fuzzy and still developing. Classification codes are not yet available for this specific complex field, because DNA-Nano's scope does not certainly fall within even one or more traditional classifications such as nanotechnology or biochemistry. Keywords that can precisely distinguish DNA-Nano from adjacent fields are hard to find.

Finally, we adopted an approach that we learned through intensive interaction with technology and business intelligence units in the industry that work on patent landscaping and patent text-mining. Unsatisfied with traditional patent selection methods (specifically based on keywords and IPC codes), these industry experts pioneered their own approach and found it useful for capturing patents in emerging fields. To the best of our knowledge, the method they developed is new to scholarly studies, and we will refer to it as the "concept approach." In short, it works as follows: First, one operationalizes the definition of a *knowledge field* into a minimum number of independent concepts (often 3 or 4), each representing an indispensable element of the field in question. For each concept, one performs an inclusive search, aiming at a (much) high recall rather than precision (for instance, using all known synonyms related to the concept, combined with the OR operator). As a second step, one selects only the intersection of all concept groups, resulting in a much smaller set. Precision is achieved at this second stage. The concepts approach is an iterative process, whereby the results of each step are monitored in terms of achieved levels of recall and precision,[11] and search queries are refined until no further improvement can be reached, and the sought level of recall and precision is achieved. While originally developed for patents, this approach can be equally used for publication retrieval.

We applied this concept approach on the knowledge field of "DNA Nanotechnology" (terminology often used in both publications and patents), and "DNA Nanoscience," by which we mean the scientific domain of DNA Nanotechnology (see Douglas, 2016, for a more elaborate discussion on the concept of DNA Nanoscience). The journal Nature Research (2018) defines DNA Nanotechnology as "*a branch of nanotechnology concerned with the design, study and application of*

[11] Precision can be estimated by taking a random sample of the set, and manually investigating whether all the records indeed belong to the sought field. Recall can be estimated by independently creating a set of records that are known to belong to the sought set (e.g., by asking an independent expert in the field, or selecting the relevant patents or publications of key contributors) and then testing whether these records are present in the set.

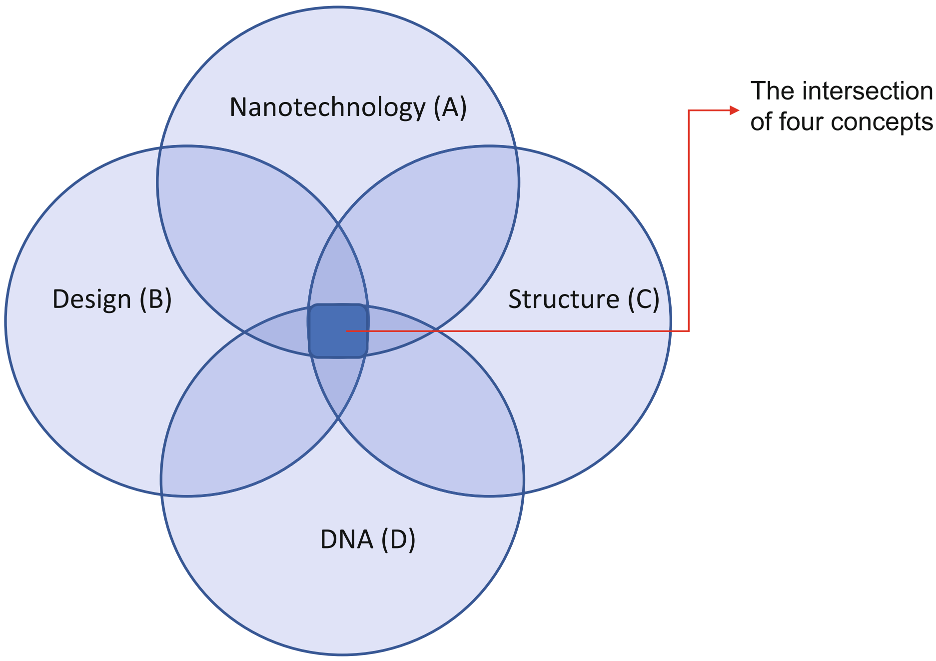

Fig. 1 Illustrating the concept approach to DNA nanotechnology

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 a literature review and on consultation with active researchers in DNA-Nano we met at conferences, we derived four[12] independent concepts to use in our concept approach. These are Nanotechnology (A), Design (B), Structure (C), and DNA (D), as illustrated in Fig. 1 (see also Annex). For each concept, we developed search queries that used all relevant keywords and known synonyms, which were collected exhaustively from multiple sources.[13] Ideally, we want to apply the same query for publications and patents, as being described in our previous work (La & Bekkers, 2018). However, investigating the relevant publications and patents of known scientists and inventors in this field, we learned that the language in publications is different from that in patents. The language in publications tends to be broader, while the language of patents is narrower and more precise. Consequently, we had to adapt our queries to the different language use in publications and patents, in order to achieve both high recall and satisfactory precision. Consequently, we employed a set of queries to collect publications, and another set of queries to

[12] We found that, in our context, four was the number of concepts allowing us to reach the best balance between recall and precision. With three concepts, the level of precision reduced significantly. With five concepts, the concepts started to lose their initial independence, and the level of recall dropped.

[13] Information sources include materials and notes taken at technical conferences on DNA-Nano, communication with experts by email and Skype, and publications and news items in the field of DNA-Nano.

collect patents. We involved two experts[14] to validate that queries included appropriate keywords. Subsequently, for each dataset, we selected the records satisfying all four concept groups. To improve the precision of each dataset, we imposed two lists of exclusion terms, one to remove the irrelevant records from the titles, and the other to remove irrelevant records from titles, abstracts, and keywords. We found these exclusion terms by reading the irrelevant records retrieved from the overlap of the four concept groups. After a number of iterative steps of improvement and refinement,[15] we created our final datasets. Because the patent dataset was much smaller than the publication dataset (there are considerably fewer patents than publications in this area), we complemented the identified patent data with their forward citations. This step further increases recall, while testing confirmed there was no notable drop in precision. (The publication set was already sufficiently large, so we did not have to take such a step). Annex provides details on the concepts we used, as well as our final search queries.

3.2 *Selecting Knowledge Areas Within S&T*

An important next step was to identify distinct knowledge areas in the field of DNA-Nano. The text from the title and abstract of papers and patents offers opportunities to do so, but also poses several challenges:

1. Technical terms often consist of combinations of words, rather than a single word (Nakagawa, 2000). The field we study is not an exception to that. Single words appearing with high frequencies[16] (e.g., “DNA,” “temperature”) are insufficient to describe a new concept or authors’ main contributions. High-frequency single words can become meaningful, descriptive terms if they are combined with other single words to form compound nouns (e.g., “DNA origami,” “temperature control”). We addressed this challenge by using the automatic Term Recognition algorithm proposed by Nakagawa (2000). In this algorithm, a Term Extract score is computed based on how many compound nouns have a simple noun N included as an element. In other words, the more frequently a simple noun is integrated with other compound nouns, the higher its score. Our tokenization

[14] Sungi Kim, PhD candidate at Seoul National University, validated the queries for collecting publications. Jürgen Schmied, CEO of Gattaquant, a company working in the field of DNA Nanotechnology, validated the queries for collecting patents.

[15] We improved recall by checking whether the authors and inventors whom we know are present in our search results. If not, we included more keywords from their publications/patents. We improved precision by sampling 20 records each time and checking if any record is irrelevant. Then we identified the keywords that distinguish DNA-Nano from other fields in that record, and put them in the exclusion terms.

[16] And even those with high term frequency-inverse document frequency (tf*idf).

process considers bigrams and trigrams, as long as they appear in the Term Extract list with a score.
2. Frequently occurring compound nouns can still be non-technical or non-descriptive,[17] or may fall outside our field of interest. As no software or algorithm can solve this in a fully automated way, we addressed this challenge through 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," and "acid amine."
3. Certain terms can be written in more than one way. Techniques such as stemming (cutting ends off words, e.g., from "saying" to "say") or lemmatization (finding the original form of a word, e.g., from "said" to "say") may be helpful for some words (especially verbs), but will not work for others, such as synonyms and abbreviations. To address this challenge, we manually harmonized terms (such as grouping synonyms, abbreviations) into term groups,[18] which represent knowledge areas. For example, we harmonized "3D structure" into "three-dimensional structure," "control of temperature" into "temperature control," and "Au nanoparticle" into "gold nanoparticle."
4. We counted the document frequency[19] (the number of documents where a term occurs at least once) of extracted and harmonized terms (resulting from the above steps) in our datasets across years and periods.

3.3 Measuring S&T Relatedness

As argued above, in the literature, knowledge relatedness has mostly been discussed within the realm of one single domain—science or technology. To investigate the evolving knowledge base of S&T related to a specific new field, we believe it is important to develop cross-domain measures. When analyzing S&T as two separate text corpora, one would not have to describe the interaction between them via conventional channels such as NPL references, PPPs. In this chapter, we use the S&T relatedness as a proxy for S&T interaction. More specifically, we need to clarify different types/indicators of knowledge relatedness as proxies for the extent and content of the knowledge interaction between the two domains.

Because we follow the approach of breaking down each of the two domains into smaller units—knowledge areas represented by terms, we will first discuss four indicators of cross-domain relatedness at the level of knowledge area[20]: similarity, commonality, complementarity, difference. *Knowledge similarity,* the most stringent measure of cross-domain relatedness*, occurs when the same narrowly defined*

[17] For instance, "this study," "this invention."

[18] We ended up with 109 cross-domain term groups, which have been harmonized from 400 technical terms extracted with highest scores by the automatic Term Recognition algorithm.

[19] We used Higuchi Koichi's KH coder text-mining software (Version 3a12d).

[20] A sub-domain unit of analysis.

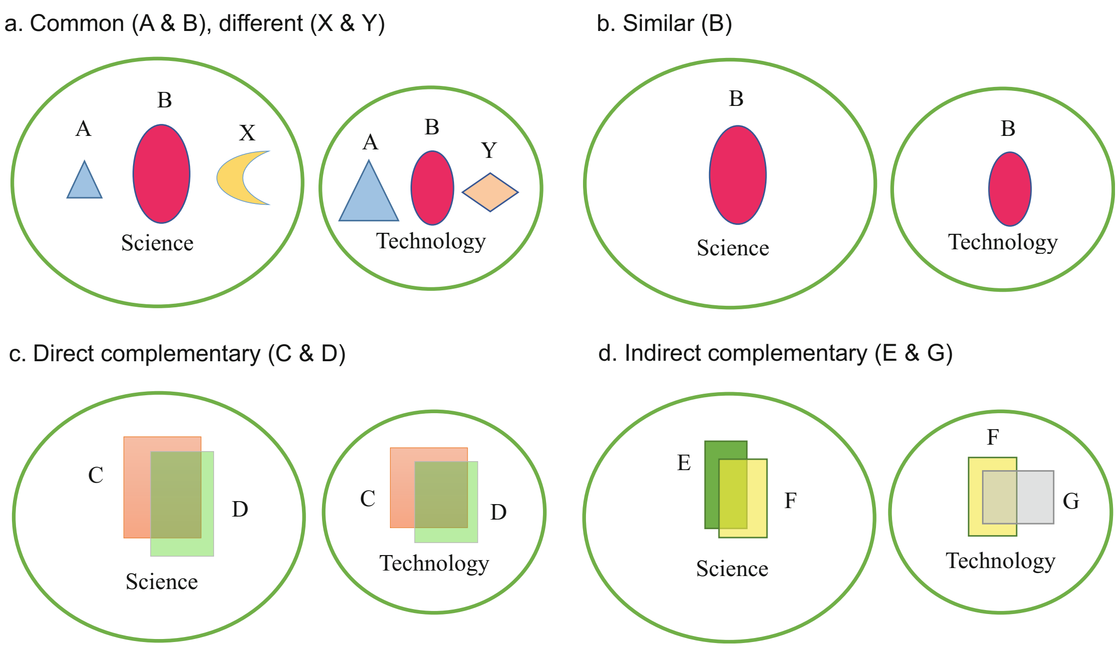

Fig. 2 Types of S&T content relatedness: commonality, similarity, complementarity (direct and indirect), and difference. (**a**) Common (A&B), different (X&Y). (**b**) Similar (B). (**c**) Direct complementary (C&D). (**d**) Indirect complementary (E&G)

knowledge area appears in both domains with similar relative frequency (example B in Fig. 2b). Similar knowledge areas are per definition common ones, but not the other way around. *Knowledge commonality occurs when the same narrowly defined knowledge area appears in both domains regardless of their relative frequency in each domain* (examples A and B in Fig. 2a). It means that the same knowledge area is used in both S&T, even if the extent of use is different. While *knowledge similarity* may indicate the highest intensity of S&T interaction, *knowledge commonality* may indicate it at a somewhat lower level. However, this is potentially useful, as a pair of common knowledge areas like C and D co-occurring in both publications and patents could strengthen the knowledge base of both S&T; a common knowledge area like F can help to bridge S&T in the case of indirect complementarity between E and G (Fig. 2d).

We furthermore distinguish two forms of knowledge complementarity in the absence of knowledge similarity. We talk of *direct knowledge* complementarity when two knowledge areas strongly co-occur in both S&T (in Fig. 2c, C and D are directly complementary). In this case, C and D are certainly common knowledge areas. However, they indicate a weaker intensity of knowledge flows between the two domains. It means that this combination frequently occurs in publications but also in patents. This should reflect the combinatory nature of each domain in an evolutionary vein. *In this case, technology relatedness coincides with science relatedness.*

In addition, we theorize *indirect knowledge complementarity* between two knowledge areas, when each of them co-occurs strongly with a third knowledge area, called a bridging knowledge area, which appears in both domains (in Fig. 2d, E and G are directly complementary, and F is the bridging area connecting them). Identifying and promoting bridging knowledge areas could help to stimulate the continuous progress of both domains economically.

Finally, knowledge areas are *different* if they only exist in one domain, not in the other (examples X and Y in Fig. 2a). This case indicates the absence of relatedness between two domains.

The above definitions relate to the individual term level. To compare two domains, the result needs to be aggregated to the domain level. We did so for the full time period of the sample, but also for three subperiods separately (see Sect. 4.2). Regarding *knowledge commonality*, we tried to identify all distinct knowledge areas (represented by terms) that two domains have in common in different subperiods, regardless of their extent. To measure *knowledge similarity*, we aimed to check if those common knowledge areas appear at a closely similar relative 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 domain (for this test, see Evert, 2005; Kilgarriff, 2001).[21] We considered the domains to be similar (i.e., belonging to some larger population) in respect of each term if the outcome is significant at 5% confidence level.

To our knowledge, no standard cross-domain measure of either *direct* or *indirect complementarity* exists. So, we propose two tests that can, in principle, be applied to any two knowledge domains. Both tests are based on the co-occurrences of terms. The first test measures the direct complementarity between two knowledge areas (represented by two terms). It is calculated as follows:

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

where J_i is the Jaccard index of the co-occurrence of the two terms in the Science domain, and J_j is the Jaccard index of the co-occurrence of the two terms in the Technology domain. Thus, our measure of *direct complementarity* J_{direct} is high when the terms in question frequently co-occur in both domains. Our second test measures *indirect complementarity* between two knowledge areas (represented by two terms). It derives from the co-occurrences of the two terms of interest with a third term, the bridging term. It is calculated as follows:

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

where J_{im} is the Jaccard index of the co-occurrence of the first term and the bridging term in the Science domain, and J_{jn} is the Jaccard index of the second term and the bridging term in the Technology domain.

[21] We used Stephan Evert's R package "corpora" for this specific Chi-square test.

Our empirical exercise for both types of knowledge complementarity involves three steps. Firstly, we reduced the co-occurrence networks of 109 terms[22] to smaller networks with only edges with a Jaccard index greater than 0.01.[23] Secondly, we matched common pairs between S&T, calculated the Joint Jaccard index,[24] then sorted and compared the lists of direct and indirect complementarity. Thirdly, we discussed our results with experts in the field (see Sect. 4.2).

3.4 Measuring the Temporal Relatedness Between S&T

Our second research question concerns the measurement of the temporal distance/relatedness between S&T. Our basic assumption here is that in modern age, what emerges at approximately the same time could be strongly related to each other.[25] We traced our list of knowledge areas, represented by the most significant terms to check if the time lag is insignificant (less than 5 years) or significant (more or equal to 5 years). We base our 5-year-threshold on the observations of Daim et al. (2007) and Finardi (2011) that a usual time lag between S&T is 3–4 years. A short time lag implies a high degree of temporal S&T relatedness. When the time lag is long, it suggests a low degree of S&T relatedness.

Note that we do not aim to determine causality here, but rather a measure of temporal relatedness. Those terms appear simultaneously in S&T could reflect the similarity between S&T, or the highest level of interaction between S&T. An inventor can file a patent first and submit a publication on the same matter right afterward. Or, scientists doing experiments in the same lab might share their collegues' work. As long as one's contribution is published or filed as a patent, other teammates can cite that contribution right away. Moreover, terms that appear with a short time lag across the S&T domains could show complementarity. There might be a hidden knowledge area in the other domain, which triggers the use of focal knowledge in one domain. In contrast, those terms appear at a longer time lag could reflect difference. In the end, we will compare with the results of our earlier analysis.

For each individual knowledge area (as represented by a term), we determined the moment it first appears (emerges) in the science domain, and when it first appears in the technology domain. While our time lag threshold of 5 years is by definition somewhat arbitrary, we believe it is appropriate to the distinction we aim to make.

[22] We explained how we selected 109 term groups in Sect. 3.2. For the actual analysis of S&T relatedness, we called them "terms" for convenience.

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

[24] This second step resulted in 133 pairs of direct complementarity and 10,525 pairs of indirect complementarity.

[25] In earlier ages, however, the temporal relatedness between S&T could happen in 2000 years (Johns, 2020).

Moreover, we carried out robustness checks which showed that variations to this threshold do not lead to substantially different results.[26]

To determine the exact moment of a term emerging the science or the technology domain, we consider the year of publication and the patent filing year, respectively. However, we aim to prevent our determination of these moments from being merely driven by an early, single, and isolated occurrence of that term. Therefore, we want to observe a certain critical mass, reflecting that knowledge has started to develop in the domain in question, rather than a one-time or incidental use of the term. For that reason, we applied a threshold: we consider the emergence of a term to be when that term hits 5% of its cumulative frequency over the full period. For most of our terms, this 5% threshold is met at the approximate value of 100 documents. Figure 3 presents an example of the time lag and threshold we applied. In our publication dataset, the term "liquid crystal" is first mentioned in 1990. Already in the same year, it reached 5% of the total cumulative frequency in 26 years. In our patent dataset, the term does not reach the 5% threshold until 1994. Therefore, the time lag between S&T regarding this specific knowledge area is 4 years. However, based on our previously mentioned criteria, we determined the time lag in this case is insignificant.

3.5 Data

Using the search queries based on our concept approach discussed above, we created a scientific publication dataset using the Web of Science (WoS) database, and a patent dataset using the Autumn 2016 version of PATSTAT. While a title of a publication or a patent is usually a set of words carefully selected by the author, it is the abstract that often mentions the relevant concepts and the contribution of authors or inventors; therefore, our queries used the text appearing in both titles and abstracts. We found 135,055 publications and 11,226 patents, dated between 1947 and 2015. However, because the WoS data on academic publications prior to 1990 often lack abstracts, we truncated both our datasets to the period between 1990 and 2015. After removing duplicates and incomplete records (e.g., publications without titles), our final datasets comprised 123,929 publications and 10,476 patents. After applying our text-mining techniques (see Sect. 3.2), we identified 109 harmonized terms that appear either solely or simultaneously in our two corpora.

To investigate the S&T relatedness over time, we further divided this 26-year time span of data into three subperiods: Subperiod 1 from 1990 to 1997, Subperiod 2 from 1998 to 2005, and Subperiod 3 from 2006 to 2015. The breaking point between Subperiods 2 and 3 is based on a ground-breaking contribution by a Caltech researcher Paul Rothemund, published in 2006 in Nature, which by September 2021 received over 4000 citations (Rothemund, 2006). The first patent for this invention

[26] These robustness checks are available upon demand from the authors.

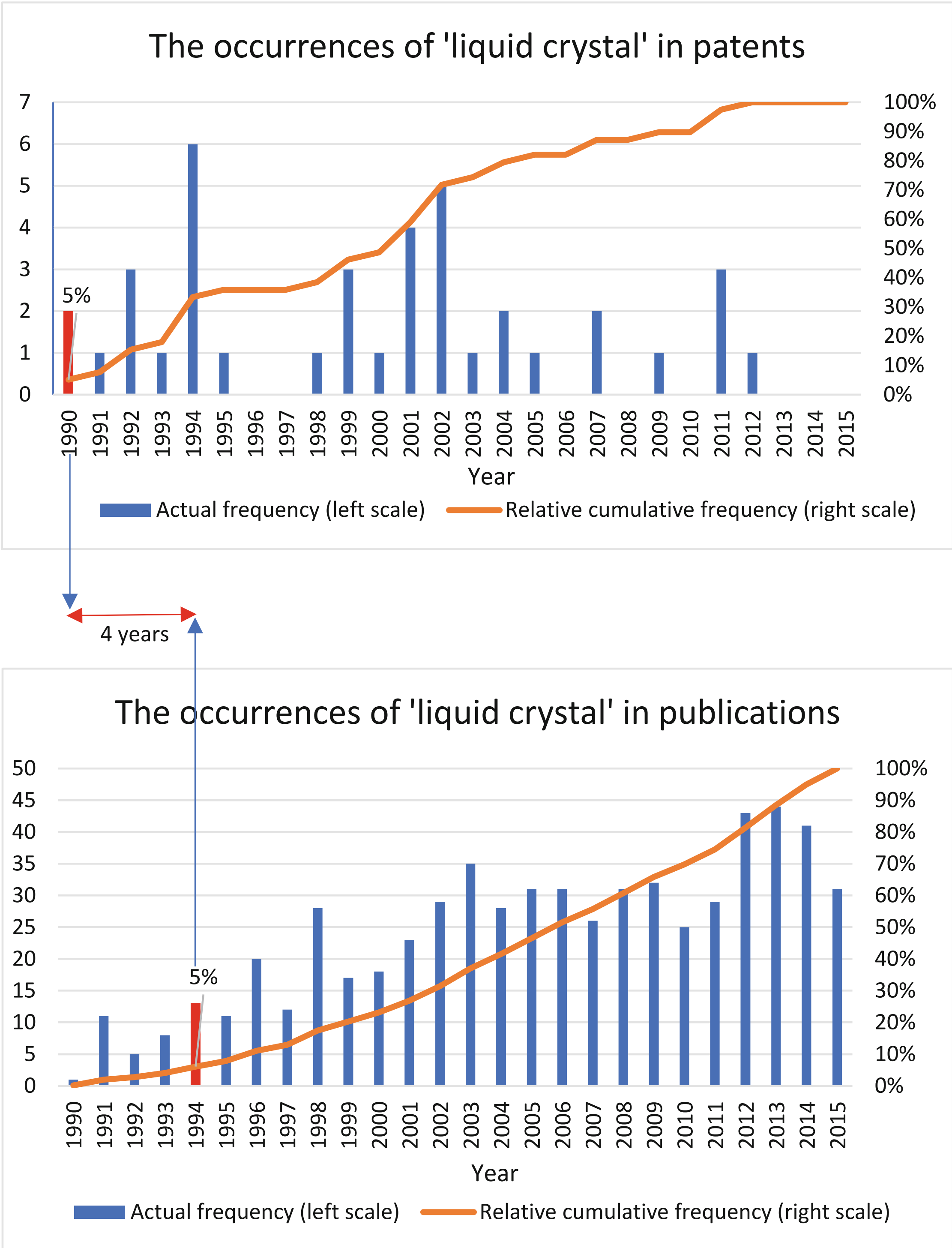

Fig. 3 The time lag of the term "liquid crystal" between S&T (represented as a red arrow)

was filed in 2005. Before and after 2006, no such compelling breaking point existed, so we chose subperiods 1 and 2 of equal length. Note that Subperiod 3 is 2 years longer than subperiods 1 and 2, which might somehow affect the data and impact the comparability. Yet, we do not expect a significant change in the number of publications, patents, and document frequencies per period due to this division. We believe that our choice of breaks between periods, based on Rothemund's breakthrough, is better than just dividing it into three equally long subperiods and ignoring this breakthrough's timing.

4 Empirical Analysis and Results

This section presents our analysis and the results we found with relation to our two main research questions (see at the end of Sect. 2.2), using the methodology described in the Sects. 3.3 and 3.4.

4.1 Descriptive Statistics of Two Corpora

Table 2 provides the descriptive statistics of our data in more details. The publication corpus is larger than the patent corpus almost 12 times in terms of the number of documents, 22 times in terms of the number of tokens, and 19 times in terms of the types of tokens. The mean of document frequency of those tokens in the publication corpus is 20, which is higher than 15 of the patent corpus. Regarding the dispersion, the standard deviation of the document frequency of the publication corpus is much higher than that of the patent corpus (371 and 81). Therefore, the publication corpus seems to be richer and more heterogeneous.

We can observe a consistent growth of the publication corpus during the full period (1990–2015), but an inconsistent growth of the patent corpus with a decline in Subperiod 3 regarding all metrics. As mentioned earlier, Paul Rothemund introduced

Table 2 Descriptive statistics of two corpora

1. Publication corpus	Full period (1990–2015)	Subperiod 1 (1990–1997)	Subperiod 2 (1998–2005)	Subperiod 3 (2006–2015)
Number of documents	123,929	22,222	37,915	63,792
Tokens in use	12,128,976	2,261,737	3,695,122	6,273,821
Types of token in use	414,234 (100%)	123,448 (100%)	178,681 (100%)	248,822 (100%)
Types of token occurring less than 5 times	359,792 (87%)	105,688 (86%)	153,728 (86%)	214,804 (86%)
Mean of document frequency	20	12	14	17
Standard deviation of document frequency	371	368	173	250
2. Patent corpus	**Full period (1990–2015)**	**Subperiod 1 (1990–1997)**	**Subperiod 2 (1998–2005)**	**Subperiod 3 (2006–2015)**
Number of documents	10,476	1679	5784	3013
Tokens in use	540,992	94,285	302,308	144,412
Types of token in use	21,741 (100%)	8125 (100%)	14,939 (100%)	9972 (100%)
Types of token occurring less than 5 times	16,414 (76%)	6221 (77%)	11,225 (75%)	7571 (76%)
Mean of document frequency	15	7	12	9
Standard deviation of document frequency	81	29	72	44

DNA origami technique in late 2005 (as a patent) and in early 2006 (as a scientific article). His contribution receives a huge number of forward citations in WoS (over 3700), but a much lower number of forward citations in PATSTAT (28). His invention is quite impactful in science, but not yet so in technology. Despite promising applications described in scientific literature, perhaps finding its way to real technological industrial applications is not so easy.

In our sample of 109 cross-domain term groups, the mean document frequency within the publication corpus is 1336, in within the patent corpus is 77. The standard deviation in the publication corpus is 1788, in the patent corpus 164.

4.2 S&T Content Relatedness

We now investigate the extent to which knowledge content in S&T domains is similar, complementary, or different, and how this evolves over time. Table 3 presents our findings, using our novel dataset and the methodologies outlined in Sect. 3.4. 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." Table 3 also shows examples of differences. For instance, in Subperiod 1, the term "microfluidic device" only appears in patents, while the term "crystal structure" only appears in publications.

Table 4 presents our findings about the degrees of commonality, similarity, complementarity, and differences in the full period and in the three subperiods. The fluctuating commonality, stable and low similarity, and increasing complementarity between the two domains suggest that the S&T domains of DNA-Nano evolve in different ways, yet achieve a higher degree of relatedness in Subperiod 3. This may be down to differences in purposes of S&T, or various knowledge recombination processes going on in each domain. Even when, using Price's analogy, these "dancers infrequently move to the same music" (low similarity), their interaction could be estimated from their increasing complementarity.

Row 1 in Table 4 presents the results of our quantitative analysis of commonality. To prevent accidental occurrences of terms in both corpora, we removed all terms with frequencies lower than five (see the discussion about common and similar knowledge areas in Sect. 3.3). This step also helps us to achieve reliable results from our Chi-square test for similarities (Rayson & Garside, 2000). We see that the degree of commonality in the whole period is high 82.6% (91 out of 109 terms appear in both domains). Looking at the subperiods, we observed that the commonality is lowest in Subperiod 1 (at 43.3%), increased in Subperiod 2 (to 80%), and started to decline in Subperiod 3 (down to 73.3%).

Row 2 in Table 4 shows the results of our similarity test for common terms that have the same relative frequency (for the Chi-square test used here, see Sect. 3.3). We found that the similarity over the whole period is only 14.4% (13 out of 90 terms have similar relative frequencies). Yet, if we consider the subperiods 1 and 3, the

Table 3 Examples of similarity, complementarity (We report pairs of complementary term in square brackets.), and difference

		Full period	Subperiod 1 (1990–1997)	Subperiod 2 (1998–2005)	Subperiod 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 crystallography, protein structure]	[Functionalization, drug delivery], [DNA sequencing, DNA fragment], [structural analysis, atomic force microscope]	[Living cell, in vivo], [structural analysis, molecular structure], [magnetic resonance imaging, magnetic properties]	[DNA origami, atomic force microscope], [gold nanoparticle, DNA origami], [self-assembly, functionalization]
3	Indirectly complementary	[Supermolecule, X-ray crystallography]	[Gold nanoparticle, carbon nanotube]	[Cancer treatment, DNA amplification]	[Cancer treatment, cancer diagnosis]
4	Bridging area	Electron microscope	Functionalization	Cancer cell	Cancer cell
5	Different (only in publications)	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, g-quadruplex DNA
6	Different (only in patents)		Microfluidic device	Hybridization chain reaction	Nucleic acid array
7	Absent in both domains		DNA origami, hybridization change reaction, short hairpin RNA, bio stability, RNA interference, semiconductor device		

Table 4 Overview of S&T relatedness indicators

	Number of terms	Full period	Subperiod 1 (1990–1997)	Subperiod 2 (1998–2005)	Subperiod 3 (2006–2015)
1	Common terms[a]	90/109 (82.6%)	39/90 (43.3%)	72/90 (80%)	66/90 (73.3%)
2	Similar terms[a]	13/90 (14.4%)	9/39 (23.1%)	14/72 (19.4%)	16/66 (24.2%)
3	Directly complementary pairs of terms[b]	133	85	127	133
4	Indirectly complementary pairs of terms	10,525	n/a	n/a	n/a
5	Different terms, only in publications	18/109	50/109	31/109	42/109
6	Different terms, only in patents	1/109	3/109	4/109	1/109
7	Absent in both domains	0/109	17/109	2/109	0/109

[a]Excluding terms with fewer than 5 occurrences
[b]Data are cumulative (up to and including the listed period)

similarity level is higher (23–34%). Subperiod 2 has lowest similarity (19.4%), which might be an indirect cause of the drop of the number of patents in Subperiod 3. The similar terms seemed to be established knowledge areas in both domains, such as liquid crystal and biological material.

For the full period, we identified 133 pairs of direct complementary terms. For the three subperiods, that number increased from 85 to 133. By definition, indirect complementary terms can occur in much higher numbers and we identified no fewer than 10,525 of these. Because of computational limitations, we did not analyze indirectly complementary terms for the different subperiods. Some similar and complementary knowledge areas form the mainstream of DNA-Nano.[27]

Table 4 also provides the result from the absolute differences (unrelatedness) between the two corpora. The level of difference is low (19 out of 109 terms are different) in the full period (as the reflection of the high commonality in the full period), is highest in Subperiod 1, and drops almost by a half in Subperiod 2 and slightly increases in Subperiod 3. The number of terms showing up only in publications is higher than terms showing up only in patents. This may trigger a thought that there are still many promising applications, which discovered by scientists but not yet materialized into real applications.

To have our findings validated by experts in the field, we asked six experts attending a major conference in DNA Nanotechnology.[28] One was Nadrian Seeman, whom we already mentioned as the founding father of this field. When we presented the similar terms we found, these experts indeed recognized them as similarities

[27]This does not happen with knowledge areas that are neither similar nor complementary.

[28]We did so at the third workshop on Functional DNA Nanotechnology (6–8 June 2018, Rome, Italy).

between S&T and believed they were the result of S&T interaction. Regarding complementarity, the experts agreed on 98% of our pairs of direct complementary terms (133).[29] However, because our list of indirectly complementary terms is so long (10,525 terms), we could neither ask the experts to check them all nor suggest any priority of importance. Perhaps future research can find ways to identify the most prominent indirect complementary terms.

4.3 Temporal Relatedness Between S&T

We measured the time difference between the emergence of knowledge areas (represented by terms) in S&T, as the proxy for S&T temporal relatedness. As we can only observe such time differences if a term appears in both domains, we excluded the 19 terms (out of 109 original terms in our datasets) that do not appear in both domains or have a frequency of only 5 documents or less. This left us with 90 terms for which we measured time lags.

From these 90 terms, 72 (80%) emerged with insignificant time lags between S&T (Group 1, examples in Box 1), which implies a strong temporal relatedness. A total of 18 terms (19.8%) emerged with significant time lags between S&T (Group 2), which implies a weak temporal relatedness: 7 emerged in science significantly earlier than in technology (Group 2a, Table 5), and 11 terms emerged in technology significantly earlier than in science (Group 2b, Table 6). These numbers could show signals of technology leads, in comparison to science.

Box 1. Examples of Terms with Insignificant Time Lags Between S&T (Group 1)

DNA origami, DNA synthesis, cancer diagnosis, self-assembly, carbon nanotube, mass spectrometry, atomic force microscope, therapeutic agent, supermolecule, RNA synthesis, DNA fragment, resonance energy transfer, temperature control

For a better understanding of types of knowledge areas emerged with a strong or weak temporal relatedness, we looked at the terms in more detail. Terms in Group 1 (e.g., DNA origami, DNA synthesis, self-assembly, etc.) represent the knowledge areas where knowledge in S&T emerged and developed almost simultaneously. Scientific and technological knowledge might originate from the same place, the same person, or be the result of a co-creation process by scientists and inventors

Table 5 shows the list of 7 terms (Group 2a), which emerged significantly earlier in science than in technology. These 7 terms represent the knowledge areas with a

[29] We explained the concepts of direct and indirect complementarity, and gave them the list of 133 pairs of terms. Some experts reacted right away, others responded later by email.

Table 5 List of terms with significant time lags (Group 2a)

	Term	When threshold was 5% total frequency of each term in publications (1)	When threshold was 5% total frequency of each term in patents (2)	Time lag between (1) and (2)
1	X-ray crystallography	1993	2004	–11
2	Crystal structure	1994	2002	–8
3	E-coli	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

Table 6 List of terms with significant time lags (Group 2b)

	Term	When threshold reached 5% total frequency of each term in publications (1)	When threshold reached 5% total frequency of each term in patents (2)	Time lag 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

weak temporal relatedness between S&T. This could happen when scientific development occurred much earlier, but only after a long period it can be realized into real manipulations/applications in technology. This is the case of "DNA structure," for which Seeman constructed the theoretical foundation, however, realized into real structures only much later. Seeman and his followers encountered many practical challenges before Rothemund stepped into this field in late 2005, early 2006. Sometimes, it could be the case that innovation development (starting from R&D projects) could not pass the valley of death, or not become successful commercially (e.g., "Raman spectroscopy"). Sometimes, it could be some knowledge areas inherited from other science fields ("X-ray crystallography," "crystal structure," "molecular biology"), but turned out not to find much use in technology.

Table 6 presents the list of terms that emerged in technology significantly earlier than in science (Group 2b). These 11 terms also represent the knowledge areas with a weak temporal relatedness between S&T. Among the 11 terms found, "hybridization chain reaction" is the one with the longest time gap between technology and science. Not only driven by techniques (DNA hybridization, DNA detection, magnetic resonance imaging) and applications (biosensor, drug delivery), technology also took the lead in "programmability" and "functionalization," which turn structures into devices. To build machines at the nanoscale, technology signaled what it needed from science: "mechanical properties."

The 11 terms in Table 6 are knowledge areas where science indeed lagged behind technology. We note that some are closely linked to medical healthcare, such as biosensor and magnetic resonance imaging. The long investment process required by firms and other actors in those areas may have resulted in patented inventions, whose diffusion to academia took time. These may be the areas where scientists needed time to recognize the relevance to their work, time to set up collaborations with industry, then use them in the context of their own research on DNA-Nano. Especially where "science of the artificial" is concerned, technology comes first in the form of workable structures, devices, and artifacts, which later become the subject of scientific research. It is also important to bear in mind that laboratory works always involve equipment, some of which may have been patented several years earlier. Traditional enabling techniques, methods used in long-existing knowledge fields (such as molecular biology and biotechnology), are still usable/ recombined in emerging DNA Nanoscience.

5 Summary, Discussion, and Conclusion

While the economic, innovation, and management literature extensively discusses knowledge development and relateness in both the science and in the technology domain, few studies look at the interaction and knowledge relatedness across these domains. This study proposes a systematic way of measuring such cross-domain S&T interaction relations. Starting from the concept of S&T relatedness (both over content and time), we introduce five novel indicators of knowledge relatedness across S&T, as shown in Table 7 (in decreasing level of S&T interaction). Following a text-mining approach, we provide the actual degrees of relatedness across DNA-Nano S&T according to the five above indicators and detect important knowledge areas across S&T, which is helpful for research evaluation, funding, and policy recommendations.

Applying our measures to the case of DNA-Nano, a research field that has delivered interesting developments in both S&T, we summarize our observations on the above indicators in Fig. 2. We find that the level of knowledge similarity, the

Table 7 Indicators for knowledge relatedness across the S&T domains

Knowledge similarity	Share of narrowly defined knowledge areas that appear in both domains with similar relative frequency
Knowledge commonality	Share of narrowly defined knowledge areas that appear in both domains regardless of their relative frequency in within each domain
Knowledge complementarity (direct)	Share of narrowly defined knowledge areas that strongly co-occur with each other in both S&T
Knowledge complementarity (indirect)	Share of narrowly defined knowledge areas that strongly co-occur with a "bridging knowledge area" that appears in both domains
Knowledge differences	Share of narrowly defined knowledge areas that exist in one domain but not in the other

most stringent measure of cross-domain relatedness, is relatively low: only 14.4% of the narrowly defined knowledge areas (represented by compound noun terms) we distinguish in our case qualify as similar. This indicator remains relatively stable over time (see Fig. 4). Yet, knowledge similarities only reflect a part of the whole picture of S&T relatedness. The commonality measure delivers a more interesting trend. Over the three subperiods, it grows from 43.3 to 80% and falls back slightly to 73.3%. Direct knowledge complementarity also goes up over time, without a fallback.[30] Differences in knowledge drop considerably from subperiod 1 to subperiod 2, but grow slightly towards subperiod 3.

The overall low degree of similarity and the increasing complementarity may indicate that although S&T interaction in this knowledge field started low, it increased and then stabilized. We may expect more industrial applications in the coming period (after Subperiod 3). Altogether, we believe this case illustrates how our proposed measures provide a sophisticated view on the development of knowledge relatedness across S&T.

While S&T similarity is the form of knowledge relatedness most discussed in the existing literature, this measurement seems mostly limited to the field's mainstream, where S&T have overlapped, intertwined, and most strongly related. Our empirical results show that we get a much more complete picture if we also measure S&T complementarity. Taking Price's analogy of a pair of dancers, S&T do not need to be identical and too close to each other. It is challenging for dancers to move if they appear to be too close to each other. While their similarities help with their sustainable and incremental movement, their complementarities encourage more knowledge recombination and learning between them. In the future, they could take more innovative, and breakthrough steps resulted from their current learning and interaction process. Without including the measure of temporal relatedness, one may not be

[30] For reasons indicated in Sect. 4.2, we did not measure indirect complementarity over the different subperiods.

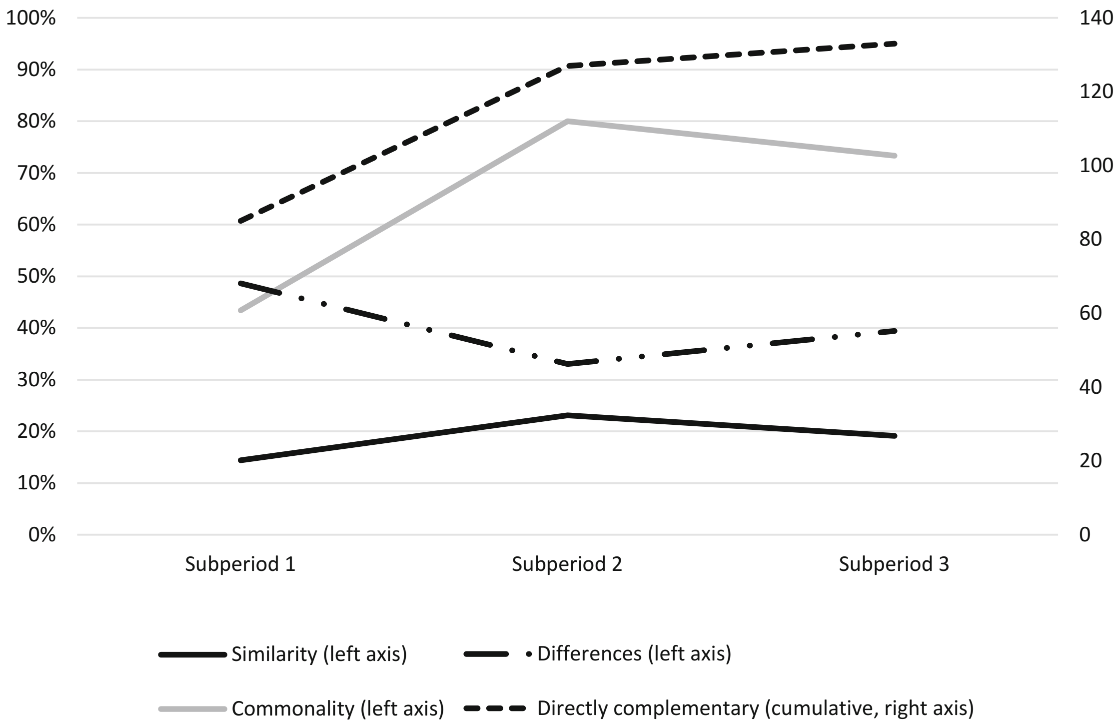

Fig. 4 Knowledge relatedness (The number of possible complimentary pairs is a very high number, and therefore the measurement of direct complementary cannot be plotted in a relative scale. See Sect. 4.2 for more details.) across the S&T domains in DNA-Nano

able to see the leading role of either science or technology, and that they grow together, hand in hand!

The S&T complementarity is our second important indicator of S&T relatedness, which has been neglected in previous empirical literature (possibly because it is much harder to measure). In practice, S&T complementarity is harder to be recognized as a form/indicator of S&T interaction. This could originate from academic-university partnerships, which facilitate knowledge exchange, equipment sharing, or star scientists' collaborations in industrial projects. Identifying complementary knowledge areas across S&T could help establish future crucial partnerships, co-authorships, co-patenting, co-location, co-creation of potential innovations, and promote technology transfer from university to industry. Funding "bridging knowledge areas," e.g., "electron microscope," "functionalization," "cancer cell," from public investment might provide a necessity for S&T's future development economically. Knowledge complementarity might reveal the combination and recombination process within each domain and the matching capabilities across these essential domains. These processes would help generate synergies, reduce R&D costs, promise the growth of more emerging science-based technologies, technology transfer in the future.

The knowledge difference indicator reflects the knowledge areas which have not yet been developed in one of the two domains. When the time lag of knowledge areas between S&T (e.g., "hybridization chain reaction") is too long, it could result in a different knowledge area. However, it is not always the case. Some knowledge areas only show up in one domain in the end of one subperiod could also appear in both domains in the early subsequent subperiod. This is the case of different knowledge areas per period but with insignificant time lags. The analysis of the relatedness per year could show a better picture of knowledge evolution. However, our statistical tests may not be implemented because of low or zero frequencies of some terms in some years, especially in the patent corpus.

The specific approach we chose for our studies also has limitations. Among other things, it does not observe a direct link between the S&T domains. Subsequent research could further explore the interaction between these domains by direct linkages such as NPL citations and PPPs and provide more insights into the knowledge recombination processes within each domain. Future studies could also apply our measures to a wider range of fields, and perhaps generate stylized facts about different types of relatedness, or a taxonomy.

Acknowledgments The Vietnamese Government sponsored this research in an overseas training program for Ph.D. candidates. We are indebted to Önder Nomaler and Cort MacLean Johns for their discussion and support. We acknowledge Higuchi Koichi for his technical support with the software KH Coder. We are thankful to have insights into the practical research of DNA Nanotechnology, presented and enlightened by experts at these conferences: DNA22, DNA meets Plasmonics, FDN2018. We are grateful to Nadrian Seeman for validating our results, and Jürgen Schmied and Sungi Kim for verifying our queries, Bas Rosier for sharing his insights on DNA Nanotechnology. We especially appreciate the constructive comments from discussants Virgilio Failla, Hans Christian Kongsted, the audience at the DRUID18 conference, from Piergiuseppe Morone at the ISS2018 conference, and two anonymous reviewers for this book chapter. We would also like to thank Nguyen Van Dan (Academy of Finance) for moral support.

Compliance with Ethical Standards

Funding This study was funded by a Vietnamese Government scholarship for conducting Ph.D. studies abroad, according to Decision No. 1908, 30 May 2013, by the Ministry of Education and Training. Mrs. Hanh Luong La was the recipient of this scholarship. She also received the travel grant from the Eindhoven University of Technology to present the earlier version of the chapter at the ISS2018 conference in Seoul.

Conflict of Interest The authors declare that they have no conflict of interest.

Annex. The Four Concepts Applied in the Concept Approach[31]

1. Description of the Four Concepts, as Well as the Exclusion Mechanisms Used

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, 2000). Thus, any science or technology that works below the scale of 100 nanometers is considered "nanotechnology." This definition is a broad one. We could therefore maximize our search by collecting synonyms referring to nanoscale or instruments 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. 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 afterward. We found all terms related to this process and listed them under the concept "design"
Concept C: Structure	Structure is defined by the Oxford Dictionary as "a particular arrangement of parts." We found several synonyms of "structure" based on publications and patents in DNA-Nano by top contributors in the field such as Nadrian Seeman, Paul Rothemund, and others. We also noted specific words related to DNA structures and included them in our 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, as used in our research, refers to artificial DNA, not its natural form. However, its

(continued)

[31] It is worthwhile noting that some records where the concept "nanotechnology" is implicit, should be included in our datasets. Certain inventors choose not to mention nano-related terms explicitly or discuss only DNA or oligonucleotides. That might be the reason why a considerable number of patents belonging to DNA Nanotechnology 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 discussing DNA or nucleotides, we should imply that the research is conducted at the nanoscale, since the dimension of a DNA strand is approximately 2.5 nm. Therefore, "DNA" and "*nucleotid*" are included in two concept areas (Nanotechnology and DNA) to avoid missing certain records that do not mention nano-related terms. Although DNA and Nanotechnology are closely related concepts, we have not grouped them because this leads to considerably more 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.

	synonyms and related terms are borrowed from molecular biology
Exclusion terms in Titles (E1)	At the first level of exclusion, we excluded specific terms relating to other closely linked fields (e.g., molecular biology, genetic engineering, forensics). However, these terms could still 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 not appear in titles, abstracts, and keywords. This strongest exclusion has improved the precision of our data

2. Final Queries

Query for Publications

(**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 = RIBONUCLEIC OR 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 '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

Query for Patents

(**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$ OR '*NUCLEOTID*' OR DNA

Design = CONJUGAT* OR FORM* OR FOLD* OR JUXTAPOS* OR PROGRAM* OR DESIGN* 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' OR ORIGAMI

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

DNA[32] ='DNA ACTUAT*' OR 'DNA NANOTECHNOLOGY' OR 'FOLDING DNA' OR 'DNA STRUCTURE' OR 'DNA ORIGAMI' OR 'DNA COMPUT*' OR 'DNA HYBRIDIZ*' OR '*NUCLEIC ACID*' OR 'DOUBLE HELI*' OR HELICES OR '*STRAND*' OR '*NUCLEOTID*' OR FOLDAMER$ OR APTAMER$

E1 = RIBONUCLEIC OR '*RNA' OR RADIOTHERAPY OR SPECIES OR '*ORGANISM' OR ORGAN$ OR 'BIOLOGICAL AGENT$' OR BIOMARKER$ OR ENHANC* OR '*BASE' OR PAIR* OR '*NUCLEOTIDE SEQUENCE' OR

[32] It is harder practically to achieve precision in retrieving patents rather than retrieving publications. Therefore, we decided to adjust terminologies in this DNA concept group into more specific terms, which include DNA.

RECEPTOR$ OR '*WEAR' OR CLONING OR BIOSENSOR$ OR SYMPTOM$ OR AGGLOMERATION OR PURIF* OR INFLAMMAT* OR 'DNA SYNTHESIS' OR HEMOGLOBIN OR HIV OR BIOACTIVE OR 'DNA AMPLIFICATION' OR 'NUCLEIC ACID AMPLIFICATION' OR BLOOD OR VITAMIN$ OR IMMUN* OR ANTIBODY OR ANTIGEN$ OR REAGENT$ OR ENCOD* OR VIRUS* OR BACTERI* OR GENE$ OR 'GENE EXPRESSION' OR HUMAN$ OR PATIENT$ OR LIFE OR 'AMINO ACID$' OR TISSUE$ OR 'NON-NUCLEIC'

E2 = 'GENE INTERACTION' OR TRANSFECT* OR TRANSLOCAT* OR PHENOTYPE OR HYDROGEN OR ENHANCER$ OR EVOLUTION* OR EMBRYO* OR SEA OR FISH* OR 'SIDE EFFECT$' OR CULTURE OR FLOWER* OR CARBOHYDRATE$ OR INHIBITOR$ OR MOUSE OR MICE OR CO-EXPRESSION OR POLYMORPHISM OR NON-CODING OR COPY OR COPIES OR PARENT$ OR EXON*

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*, 351–370.

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

Bekkers, R., & Martinelli, A. (2012). Knowledge positions in high-tech markets: Trajectories, standards, strategies and true innovators. *Technological Forecasting and Social Change, 79*, 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*, 69–82.

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

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*, 107–114.

Boyack, K. W., & Klavans, R. (2008). Measuring science–technology interaction using rare inventor–author names. *Journal of Informetrics, 2*, 173–182.

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

Chang, Y. W., Yang, H. W., & Huang, M. W. (2017). Interaction between S&T in the field of fuel cells based on patent paper analysis. *Electronic Library, 35*, 152–166.

Daim, T., Monalisa, M., Dash, P., & Brown, N. (2007). Time lag assessment between research funding and output in emerging technologies. *Foresight, 9*(4), 33–44.

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

Douglas, K. (2016). *DNA nanoscience: From prebiotic origins to emerging nanotechnology*. CRC Press.

Drexler, K. E. (2013). *Radical abundance: How a revolution in nanotechnology will change civilization*. Public Affairs.

Dunn, K. E. (2020). The business of DNA nanotechnology: Commercialization of origami and other technologies. *Molecules, 25*, 377.

Evert, S. (2005). *The statistics of word cooccurrences word pairs and collocations*. Dissertation, Institut für maschinelle Sprachverarbeitung, Universität Stuttgart. Retrieved from http://elib.uni-stuttgart.de/bitstream/11682/2573/1/Evert2005phd.pdf

Finardi, U. (2011). Time relations between scientific production and patenting of knowledge: The case of nanotechnologies. *Scientometrics, 89*, 37–50.

Heinisch, D., Nomaler, Ö., Buenstorf, G., Frenken, K., & Lintsen, H. (2016). Same place, same knowledge–same people? The geography of non-patent citations in Dutch polymer patents. *Economics of Innovation and New Technology, 25*, 553–572.

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*, 2005–2022.

Johns, C. M. (2020). *The industrial revolution-lost in antiquity-found in the renaissance*. Pumbo.nl.

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

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

La, H. L., & Bekkers R. N. A. (2018, June 11–13). *The relation between scientific and technological knowledge in emerging fields: Evidence from DNA nanoscience and DNA nanotechnology*. Paper presented at DRUID18 conference, Copenhagen, Denmark.

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

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*, 602–628.

Meyer, M. (2000). Does science push technology? Patents citing scientific literature. *Research Policy, 29*, 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*, 789–806.

Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: Exploring tissue engineering. *Research Policy, 31*, 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*, 195–210.

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

Nature Research. (2018, January 1). DNA nanotechnology. Retrieved from https://www.nature.com/subjects/dna-nanotechnology

Nelson, R. R., & Winter, S. G. (1974). Neoclassical vs. evolutionary theories of economic growth: Critique and prospectus. *Economic Journal, 84*, 886–905.

Nomaler, O., & Verspagen, B. (2008). Knowledge flows, patent citations and the impact of science on technology. *Economic Systems Research, 20*(4), 339–366.

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: Principles and fundamentals* (pp. 217–244). Wiley.

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

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

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

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

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

Rayson, P., & Garside, R. (2000). Comparing corpora using frequency profiling. In: *Proceedings of the workshop on comparing corpora* (vol. 9, pp. 1–6). Association for Computational Linguistics.
Rothemund, P. W. K. (2006). Folding DNA to create nanoscale shapes and patterns. *Nature, 440*, 297–302.
Suenaga, K. (2015). The emergence of technological paradigms: The evolutionary process of S&T in economic development. In A. Pyka & J. Foster (Eds.), *The evolution of economic and innovation systems. Economic complexity and evolution* (pp. 211–227). Springer.
Tripodi, G., Chiaromonte, F., & Lillo, F. (2020). Knowledge and social relatedness shape research portfolio diversification. *Scientific Reports, 10*(1) Nature.
US National Nanotechnology Initiative. (2000). What is nanotechnology? Retrieved from https://www.nano.gov/nanotech-101/what/definition
Verbeek, A., Debackere, K., Luwel, M., Andries, P., Zimmermann, E., & Deleus, F. (2002). Linking science to technology: Using bibliographic references in patents to build linkage schemes. *Scientometrics, 54*, 399–420.
Verspagen, B. (2007). Mapping technological trajectories as patent citation networks: A study on the history of fuel cell research. *Advances in Complex Systems, 10*(1), 93–115.
Wang, L., & Li, Z. (2018). Knowledge transfer from science to technology—the case of nano medical device technologies. *Frontiers in Research Metrics and Analytics, 3*, 11.
Zhao, Q., & Guan, J. (2013). Love dynamics between S&T: Some evidences in nanoscience and nanotechnology. *Scientometrics, 94*, 113–132.