Chapter 7

Discussion

This chapter gives an integrated overview of this PhD research followed by an interpretation of the findings and results. It is presented a general review of our central research question, namely whether patent information provides, once processed, knowledge for a better understanding of a given technology, and the contributions of this thesis to answer it. More specifically, It will be disclosed the contribution to solve the secondary research questions of :

– What are the technological components of a *piece of technology*?
– How the technological components of a *piece of technology* evolve in time?
– How are the technological components of a *piece of technology* are inter–related?, and
– What is the technological proximity of a given *piece of technology* to other technologies.

The specific contributions of each paper forming this thesis by compendium is also presented, as well as the consistency and inter-relation of said papers.

7.1 The role of Information within Patent Publications

I started this research after several decades of working in intellectual property, in particular as patent examiner and documentation classifier in the technical fields of computer graphics – *G06T* – and speech processing – *G10L* – at the European Patent Office – *EPO* –. During my work as patent examiner and documentation classifier I was in charge of searching for the prior art and the right place to classify the patent applications and technical documentation. This experience led me to observe and handle the huge and rich information contained in the bodies formed by patent publications and their prior art, and patent classification.

Non-Patent Literature	**Patent** Literature
References Selection	
- Positive and negative citations - Homage or criticism - Substantiating claims - Authenticating data - Provide background reading - Seminal papers/background papers	- Evaluating Novelty and non-obviousness - Any documents is equally important
Documentation Sources	
- Access to publishers DBs and - Access to given Libraries	- Minimum PCT collection
Citation features	
- Lack of unique identifier - Self-citation - Multiple-authorship	- Unique identifier: the PN - No self-citations - Authorship is defined by law.

Table 7.1: NPL vs Patent Literature

Patent publications and its cited prior art have some very interesting features that, make easier its collection and exploitation with the computer in comparison to Non-Patent Literature – *NPL* – (papers, theses, books, conference proceedings, ...) and its bibliographic references.

The rationale for selecting bibliographic references in *NPL* is multiple and heterogeneous whereas in patent literature, the single reason to select prior-art is clearly defined by law as the evaluation of novelty and inventive step or non-obviousness. This fact renders every patent document equally important as far as it helps to this evaluation(see in table 7.1 the *References selection*).

The sources of documentation for *NPL* authors are heterogeneous in comparison with patent examiners (see in the table: 7.1 the *Documentation sources*). For patents, in principle, every examiner of the mayor offices are looking in the same collection to pick up the documents to cite as prior art references, the *minimum PCT collection* [WIPO (2022b)]. Additionally, in papers usually seminal and background art papers are cited whereas in patents any document is equally important as far as it helps to evaluate novelty and non-obviousness.

Furthermore, some features of the cited documents differ between *NPL* and patent publications (see in the table: 7.1 the *Citation features*). In *NPL* there is not a unique identifier. Publishers have created in the late 90s the *digital object* *identifier* – *DOI* – that is a potential solution but is still facing problems such as the limited access to the *DOI* of individuals or nontraditional publishers due to the fact that it was originally designed to facilitate electronic transactions. On the contrary, every patent has a unique identifier, the patent number formed by some characters to identify the patent office – the *country code* – and a serial number.

Self-citation is an important source of noise in the aggregation and analysis of references in *NPL* whereas in patents, the author – the inventor – and the person selecting the prior art references are different, so necessarily self-citation is absent. Finally, authorship is regulated by law in patent publications, however, in *NPL* there is not authorship regulation and the frequent presence of numerous authors is an additional source of noise.

When a patent examiner selects prior art and produces the list of citations, he is in fact linking the patent application and the list of prior art, but with the sole aim of investigating novelty, and non-obviousness. Accordingly, he is linking the technology disclosed in the patent application with the technology disclosed in the prior art citations which, in the understanding of the examiner, are the closest documents to the totality or to some fundamental aspect of the technology disclosed in the application under study. No other elements, such as homage or criticism of other author's work, substantiating hypotheses, or giving grounds for any data or choices are taking into account by the examiner to select the prior art to cite.

In conclusion, patent literature is a body formed by highly structured, homogeneous and universal documents which facilitate their collection and analysis. These features, together with the fact that every patent publication is classified within the same scheme – the *IPC* –, provide a common framework for the aggregation of patents and its prior art. This framework of classification gives a very concise and clear meaning to the analysis of patents and its prior art citations, namely proximity or relatedness between the technology disclosed in the patent and the technology disclosed in its prior art citations list. Furthermore, using this classification codes we can easily aggregate technical fields of patent publications from different geographic sources, languages and periods of time, rendering possible the analysis of the technology related to a patent or a set of patents. Additionally, the fact that the *IPC* scheme is hierarchical [WIPO (2022c)] makes possible the straightforward aggregation of technical fields at different levels of granularity [Leydesdorff (2008)].

7.2 Technological Components as Characterising Factors

Let us now define a *piece of technology* as the set of patents belonging to a technical field, owned by a tech company, or granted to a researcher or group of researchers. Let us take advantage of the easiness of exploitation of prior art citations to enrich our initial set forming a sort of super-set with both the set of patents and its prior art citations. Then, by collecting their assigned classification codes, and aggregating them according to the IPC-scheme, we are able to build a new data structure, the *technology footprint* which will reveal the technologies present within our selected *piece of technology* and their weights.

We have also defined a new graphical tool, the *technology spectrum* – *TechSpectrum* –, to visualise the new data structure. We have called this tool in this way because the expression "spectrum" indicates a "complete and continuous sequence or range of opinions, elements, physical features or items between two extreme points" [Merriam-Webster (2022)]. Thus, our spectrum visualises the complete and continuous sequence of technologies between the two extremes of the *IPC* (See Figure 7.1).

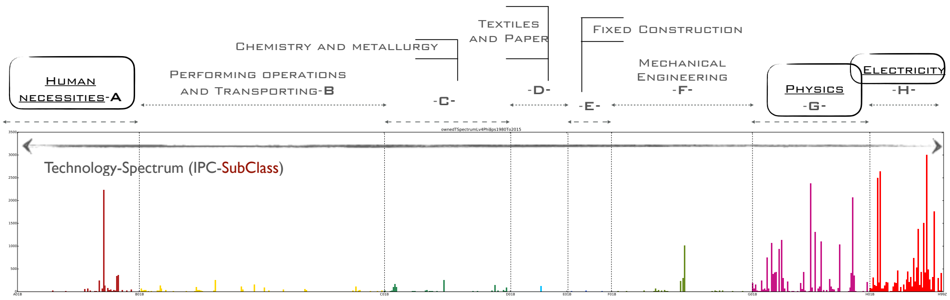

Figure 7.1: Example of *TechSpectrum* at IPC-*SubClass* level

Figure 7.1 shows a *TechSpectrum* for *Koninklijke Philips b.v.*, the Dutch transnational tech company, at *SubClass* level of the *IPC* . This *TechSpectrum* is the result of aggregating its patents from 1980 to 2015 and their corresponding prior art citations. This will be the time interval for all the cases in this thesis, in order to cover a large span of time but, of course, this time interval can be changed.

Looking at the *TechSpectrum* we can observe that the main activities of this firm are in section A – *human necessities* –, and sections G – *physics* – and H – *electricity* –. This first information is consistent with the idea of this trans-national corporation. Looking more in detail into his *TechSpectrum*, we can see that the main *technological components* are in particular in the codes *A61B* – diagnosis and surgery –, *F21V* – details in lighting devices –, *G06F* – data processing –, *G01B* – measuring instruments –, *H01J* – electric discharge tubes –, *H01L* – semiconductor devices – and *H04N* – TV – [1].

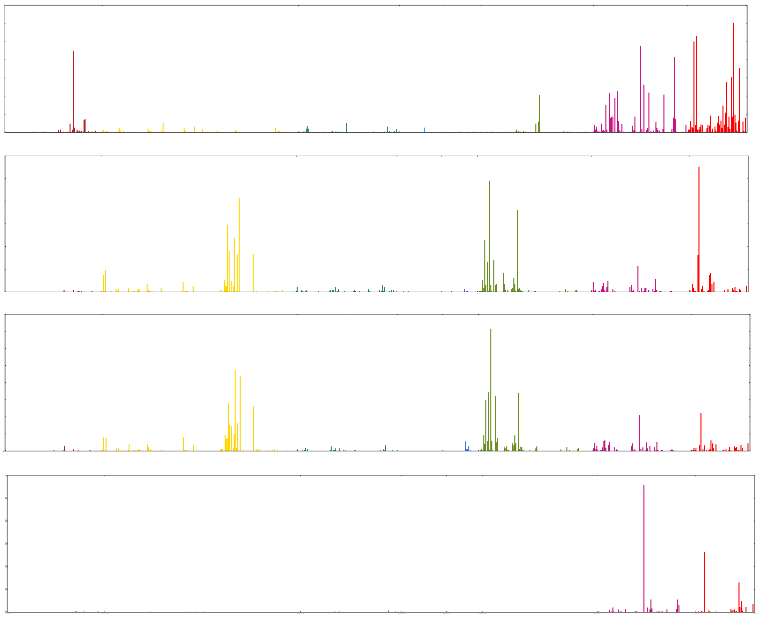

Figure 7.2: *TechSpectrums* at IPC-*SubClass* level for (from high to low) Philips, Toyota, Ford Motors and IBM

Another example can be seen in Figure 7.2 which shows the *TechSpectrum* of the following four tech companies at *IPC-SubClass* level, from top to down: *Koninklijke Philips*, *Toyota Motors*, *Ford Motors* and *IBM*. Note the remarkable similarities of the yellow[2] and the green areas[3] between the middle-top (*Toyota motors*) and middle-down (*Ford Motors*) *TechSpectrums*. This similarity indicates that both companies have very similar research and development activities in these areas of technology. This is indeed the case because both are motor companies. It is also interesting to note that the top- and the bottom-graphs represents both tech companies in the electricity-physics domain. However the top-graph (*Koninklijke Philips*) presents a broad range of developments, and, the down-graph (*IBM*) shows an extremely specialised *TechSpectrum* with virtually just two *technological*

[1]For the official titles of IPC codes see at https://www.wipo.int/classifications/ipc/en/

[2]Yellow components correspond to IPC Section B: Performing operations; Transporting

[3]Green components correspond to IPC Section F: Mechanical Engineering; Lighting; Heating; Weapons; Blasting

components at this level, namely computers – *G06F* – and semiconductors – *H01L*. –.

By analysing the *technology footprint* or observing his *technology spectrum* lot of information about the *piece of technology* nature of the developed by a company can be grabbed. Just observing the *technology spectrum* of *IBM* in Figure 7.2 we understand that it is a computer company with lots of interest in semiconductor developments. On the other hand, just observing the *technology spectrum* of *Koninklijke Philips* we will figure out that it is a medical-electronics-electricity company, very specialised in medical technology but "generalist" in electronics and electricity. In order to refine our perception of all these tech companies, we need to go deeper into the classification.

The hierarchical nature of the *IPC* scheme allows the straightforward generation of our new data structure, the *technology footprint* and the graphical tool, the *TechSpectrum*, at several levels of resolution (See section 2.2). We have chosen to use only the first four levels of the IPC, namely *Section*, *Class*, *SubClass* and *Group* because it is difficult to maintain consistency in time, and between examiners and offices for such a high level of detail as the IPC *SubGroup*. Notice that, at that level of classification resolution, the *IPC* has more than 70000 codes.

In Figure 7.3 the *TechSpectrum* of *Toyota Motors* is displayed at four levels of classification resolution of the *IPC* (from high to low): *IPC-Section*, *IPC-Class*, *IPC-SubCLass* and *IPC-Group*.

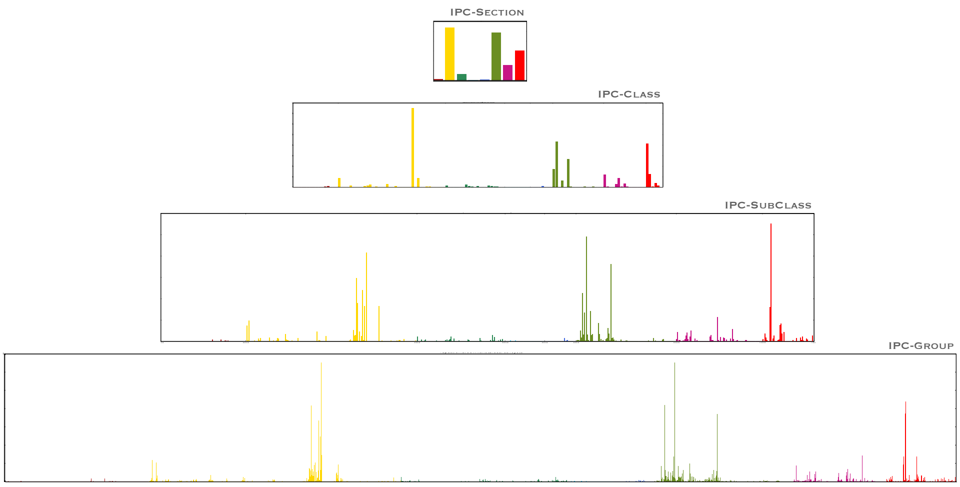

Figure 7.3: *TechSpectrums* of *Toyota Motors* at four IPC-levels, from high to low: *IPC-Section*, *IPC-Class*, *IPC-SubCLass* and *IPC-Group*

Note that in Figure 7.3, in order to facilitate the easiness of reading, the actual proportions in (horizontal) size between the different *TechSpectrum* levels have not been preserved. In fact, the first level, the *IPC-Section* has 8 codes whereas the *IPC-Class*, *IPC-SubCLass* and *IPC-Group* have 131, 647 and 7545 codes respectively. So, it is impossible to display the *TechSpectrums*, at the four levels, in the same figure maintaining the actual proportions because the first two levels would collapse in just a vertical line.

The different levels of conceptual resolution help us to form, a more refined and complete characterisation of the *piece of technology* under study. Observing the top graph in Figure 7.3 (*IPC-Section* level), two *technological components*, *B* and *F*, appear as the most and equally important components, they correspond to "Performing operations, transporting" and "Mechanical Engineering; Lighting ; Heating ; Weapons ; Blasting" respectively. At the same time, two other *technological components* stand out, although secondary: *H* and, to a lesser extent *G*, corresponding to "Electricity" and "Physics" respectively.

These components are too generic to give us a proper perception of a tech company as *Toyota motors* although, obviously we get that it is not a chemistry company, a medical technology company or a pure electronics company. Going into the following level of classification, the *IPC-Class* level, we will have a more refined picture of the company (See the second graph from the top in Figure 7.3). The main component, the yellow component, is the *B60* code, corresponding to "vehicles in general", this major component clearly characterise *Toyota motors* as a vehicle technology company. Furthermore, at this level we observe another interesting fact, the secondary components at this level are the codes: *F02*, *H01* and to a lesser extent but nevertheless strong, *F16*, corresponding to combustion engines, basic electric elements and engineering elements respectively. With all this information in hand, *Toyota motors* is clearly characterised as what it is, a vehicle, mechanics and electric technology company. This characterisation as an electric company attract our attention because, at first look, we would expect from *Toyota motors* to be a pure mechanics company. To investigate this electricity component we go again into a deeper level of classification resolution, namely the *IPC-SubClass* level visualised in the third from-the-top graph in Figure 7.3. At this level, the red component is so strong as the yellow (vehicle) and green (mechanics) components. The red one is *H01M*, the yellow is *B60W* and the green one is *F16H*, which correspond to battery technology, control systems for hybrid vehicles and gearing. Now we have a clear picture of *Toyota motors* as a company developing technology in batteries, hybrid control and gearing. To finish off the company characterisation on battery technology, just going to our last level of resolution, the *IPC-Group* level, visualised in the bottom graph in Figure 7.3, reveals that *Toyota motors* is developing in particular the manufacture of battery electrodes and the manufacture of fuel and secondary cells, these are the codes

H01M4, *H01M8* and *H01M10* respectively.

7.3 Evolutionary Dynamics of Technological Components

As mentioned in the introduction, section 1.1.1, we have called our new tool *technology spectrum* to symbolise that each of these graphs shows at its classification level the complete and continuous sequence of existing technologies between the two formal extremes of the *IPC*. We went further in mimicking the spectral representation of sound waves by defining a new data structure *Dynamic Technology Footprint* and its graphical representation, the *Technology Spectrogram* – or *TechSpectrogram* –.

In the same way that sound spectrograms are a sequence of spectrums along a time interval, our new data structure and graphical tool are built as a sequence of *technology footprints* and *technology spectrums* along a time interval, which we have defined in this thesis for practical purposes from 1980 to 2015. The time base for the individual *technology spectrums* is one year time, in the same way that time interval, this time base could be changed but we think that one year time is a good balance between aggregation and technology change.

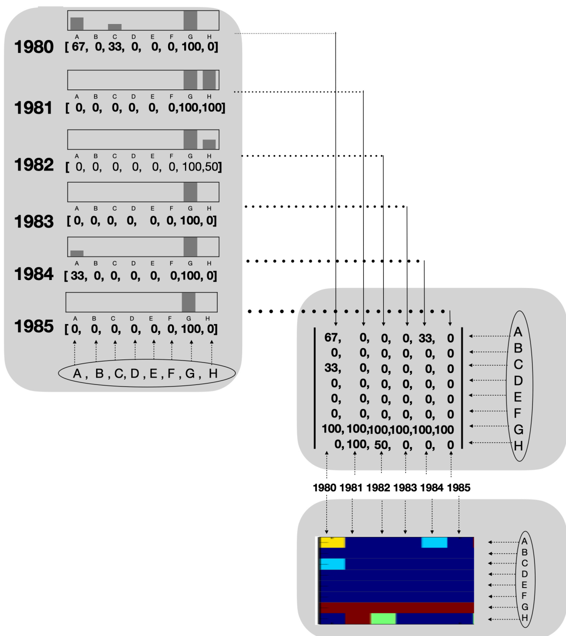

Figure 7.4: Diagram of *TechSpectrogram* formation at level *IPC-Section* from several consecutive yearly *TechSpectrums*

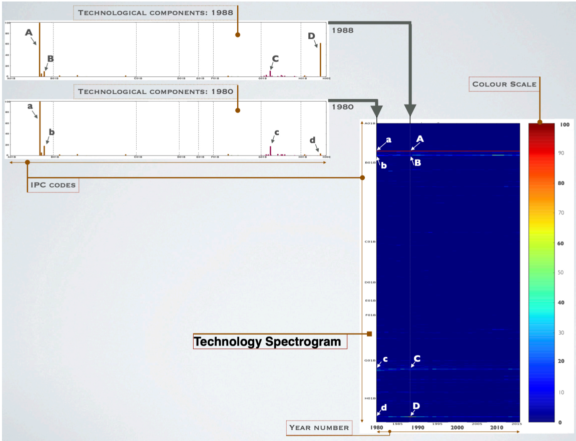

Figure 7.5: Diagram of *TechSpectrogram* formation at level *IPC-SubClass* Showing two yearly *TechSpectrums*

In conclusion, our *technology spectrogram* plots the time in the X-axis, on a year-by-year basis, and the aggregated value for each IPC-code, the *technological components*, in the Y-axis. The different colours represent the aggregated value of each *IPC* code or the strength of each *technological component*. We have represented in red the maximum strength and in blue the minimum, so zero or no existence of this component (See Figure 7.4 and Figure 7.5).

The path diagram for the formation of the first years of a *dynamic technology footprint* and its corresponding *technology spectrogram* from the first six annual *technology spectrums* is visualised in Figure 7.4. As it is shown in this figure, each yearly *technology footprint* is placed in the column corresponding to its year. The elements of each column correspond to the classification codes ordered (from high to low) according to the *IPC*, in other words, the first codes will be placed in the high positions of the columns whereas the last codes will be placed in the lower positions of the columns. Note in Figure 7.4 how, for the first yearly data, corresponding to the year 1980, the value of the first *IPC-Section*, section *A*, that is a value of 67, is placed in the first position of the first column, the value of the second *IPC-Section*, section *B*, that is a value of 0, is placed in the second position of the first column , and so on until the last *IPC-Section*, section *H*, which is placed in the lowest position of the column. Then, for the second yearly data, corresponding to the year 1981, the value of the first *IPC-Section*, section *A*, that is a value of 0, is placed in the first position of the second column, and so forth until the last year of our time interval, 2015, and the last *IPC-Section*, section *H*.

Furthermore, Figure 7.5 shows another diagram of a *technology spectrogram* formation, this figure shows how two *TechSpectrums*, in this figure from the years

1980 and 1988, are related to the formed *TechSpectrogram*. As can be seen in this figure, each *TechSpectrums* is located in the X-coordinate corresponding to the year of generation of the spectrum, the first at the X value corresponding to 1988, and the second at the X-value corresponding to 1988. The Y-coordinates are assigned from high to low corresponding to A-section codes to H-section codes.

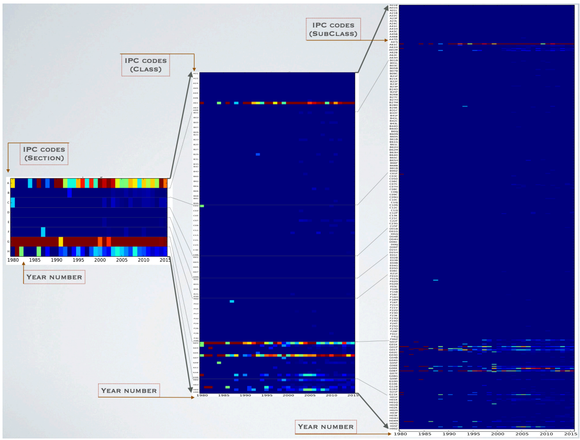

Figure 7.6: *TechSpectrograms* of *Computer Tomography* technical field at the first three *IPC* levels (from right to left): *IPC-Section*, *IPC-Class*, *IPC-SubCLass*

In a similar way that it is done for *Technology Spectrums* we have used the different levels of classification resolution of the *IPC* to generate the corresponding *Dynamic Technology Footprints* and *technology spectrograms*. The different levels of resolution will provide information, in this case, on the dynamics at each level of detail. Figure 7.6 shows the *technology spectrogram* of the *Computer Tomography* technical field – *C.T.* – at three *IPC* levels (*IPC-Section*, *IPC-Class* and *IPC-SubCLass*). As with the *Technology spectrums*, this figure does not show the three *TechSpectrograms* in its actual proportions to facilitate the interpretation.

Note how the different resolutions provide interesting information to understand the dynamic of the visualised technology. Looking at the first *TechSpectrogram*, corresponding to the *IPC-Section*, the first thing to note is that, although *C.T.* is mainly a medical device and technique, the main *Technological component* is in the *IPC-Section* G – *Physics* –. This is due to the fact that, as a technical field, it is a combination of X-ray generation and computing, and in consequence these two major techniques are reflected in its *TechSpectrogram*. Analysing the second *TechSpectrogram* (middle-graph), corresponding to the *IPC-Class*, it seems clear that for *C.T.* technology, the second main *technology component*, located at the lower part of the *TechSpectrogram*, was very active until 2000 and then the activity declined. That is noted by the red colour of the spectrogram until 1995, and the light blueish and yellowish colour afterwards. This component corresponds to measuring and testing techniques – *G01* –. On the other hand, another main *technology component*, *G06* corresponding to computation, is located below *G01*, and is very active all around the interval of observation. Looking now a step further in the *IPC* hierarchy, the *TechSpectrogram* at level *IPC-SubClass*, indicates that the *G01* activity is created specifically for measurement of nuclear or X-radiation – *G01T* –. Whereas, the *G06* activity is related to computer graphics – *G06T* – and pattern recognition – *G06K* –.

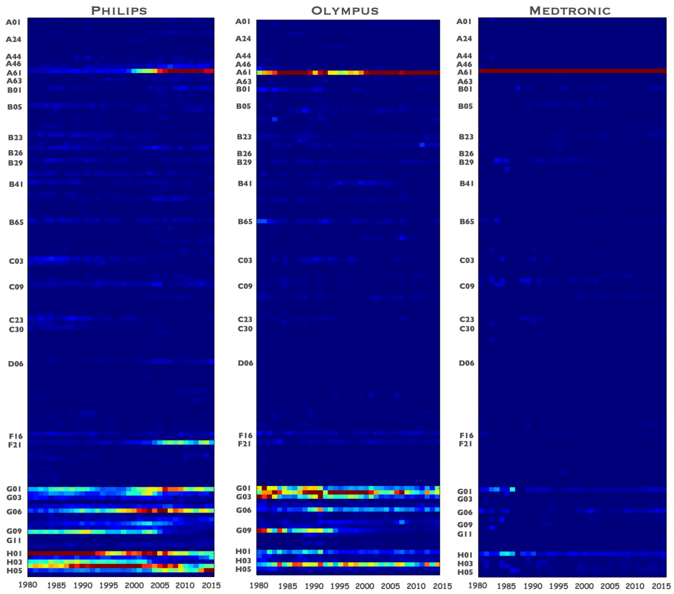

Figure 7.7: *TechSpectrograms* of three tech companies at *IPC-Class* level (from right to left): *Medtronic*, *Olympus*, *Philips*

The *TechSpectrogram* provides a rich and easy-to-grasp information about the real research and development activities of tech companies along the time interval of observation, this will improve our understanding of the *piece of technology* represented, and therefore to characterise it, to observe how the *technological components* evolve. For example, looking at Figure 7.7 it is clear that the three companies have activities in the medical field because the three of them have a *tech component* at the upper part of the spectrogram corresponding to *medical or veterinary science* – *A61* – *IPC-Class*. However, their profiles are very different. *Medtronic* appears as a firm specialised in medical technology because its *technology components* are concentrated in the *A61* component, and this is always in red, whereas Olympus appears very active in medical technology but with an important activity in *measuring* – *G01* –, *optics* – *G02* – and *photography* – *G03* –. Meanwhile, *Philips* presents a *TechSpectrogram* with broader research and development activities in *Physics* and *Electricity* fields. Note, and on the other hand, the notable increase in medical technologies from the 2000's.

7.4 Means for Exploring the Technological Components

The formal similarity of our tool with the frequency spectrogram analysis of acoustic waves [4] has deep implications because it brings the potentiality of straightforward use of the whole range of processing available for frequency spectrograms to our tool, and in consequence to the analysis of patent information.

Just as an indication of the potential of the direct use (or translation) of processing from frequency spectrograms we can see, for example, the characterisation of sound recordings by frequency spectrogram analysis applying landscape ecology techniques disclosed by Villanueva et al. [2011]. They characterise a sound recording by computing in time three parameters of the spectrogram, namely *Band Diversity*, *Band Evenness* and *Band Dominance*. This characterisation can straightforwardly be translated to our tool in order to characterise technology generating the three equivalent parameters, namely *Classification Diversity*, *Classification Evenness* and *Classification Dominance*. Of course, these three parameters can always be applied to the classification data for statistical analysis such as it was partially done by Leydesdorff [2018] but our tool facilitates this characterisation because once the frequency spectrogram is available, it requires a mere translation to be used by our tool.

We would also like to highlight the versatility of our *TechSpectrogram* at graphical level. In order to improve their interpretation we have applied some image processing directly to the *technology spectrogram* pictures, such as image enhancement (B&W conversion followed by blurring effect, and dark-blue Cutoff effect) to highlight the areas of activity – technical developments – of the spectrogram (See Figures 7.8 and 7.9).

For example, Figure 7.8 shows the *technology spectrogram* of the Japanese tech company Omron Corp. As can be seen in its *technology spectrogram*, this company is active in numerous technical fields. We have applied first a black-and-white conversion to the original spectrogram picture, and then we have applied a blurring filter to the image. The resulting image (see right-hand spectrogram in

[4]Our tool orders the complete range of classification codes sequentially in time in a similar way that the frequency spectrogram order the complete range of frequencies sequentially in time, and the classification bins play the role of the frequency bins in the acoustic frequency spectrograms

Figure 7.8) provides much more clear information about the technical areas and time intervals of activity, the interpretation of this *technology spectrogram* has been greatly simplified. At the same time, the spectrogram has been enhanced in a way that low intensity activities, pixel areas in not so dark blue in the spectrogram, are clearer. If we compare the bluish central area of the spectrogram with the same area in the filtered picture, in Figure 7.8 the right-hand graph and the left-hand side graph respectively, it is clear that in the filtered version these activities are much easier to be perceived.

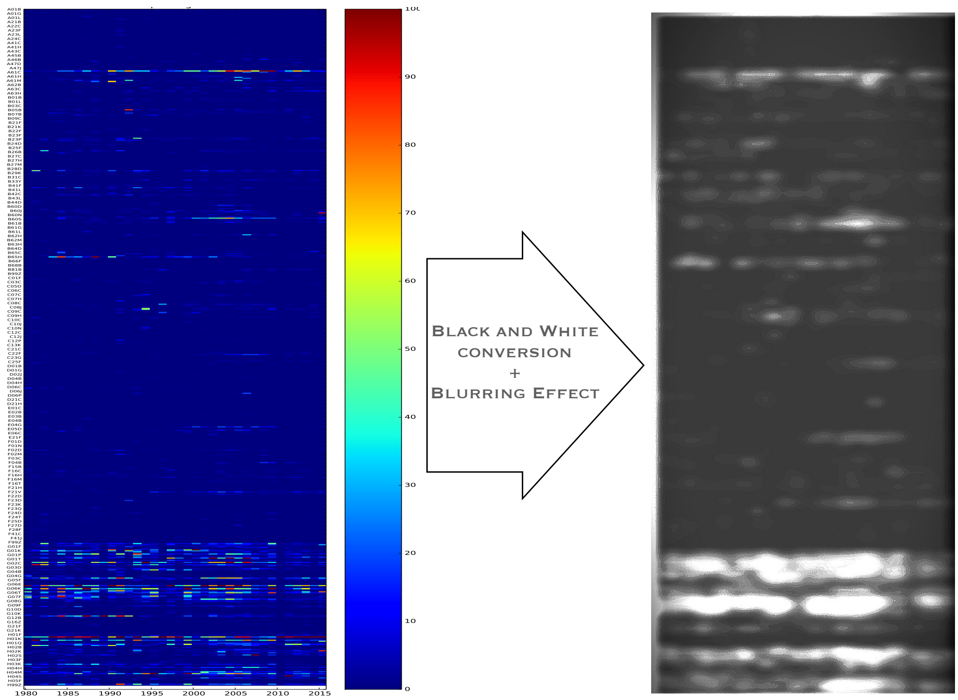

Figure 7.8: *TechSpectrogram* of Omron corp. at *IPC-SubClass*, and its B&W blurred version (right-hand graph)

Another example of image processing applied to the *technology spectrogram* can be seen in Figure 7.9. In this case, the spectrogram picture corresponds to IBM at *IPC-SubClass* classification level. The generated spectrogram has been filtered by cutting the dark bleu of the original picture [5]. The original and the filtered pictures are shown in Figure 7.9, in the right- and left-hand side graphs respectively. The filtered spectrogram enhanced the picture by removing all information except the areas above a 25% of activity. The filtered picture, it is easier to read if we are focusing our interest in the areas above a certain activity, in this example 25%,

[5] The Cutoff threshold is set at 25% of activity

and therefore areas with pixels from light-blue to red colours. As in the previous example, the filtered pictures are easier to segment and automatically identify the areas of interest. The straightforward use of image processing packages on our graphical tool represents a great potential. This is not the case, to the best of our knowledge, for much more complex visualisations based on patent information such as patent landscapes.

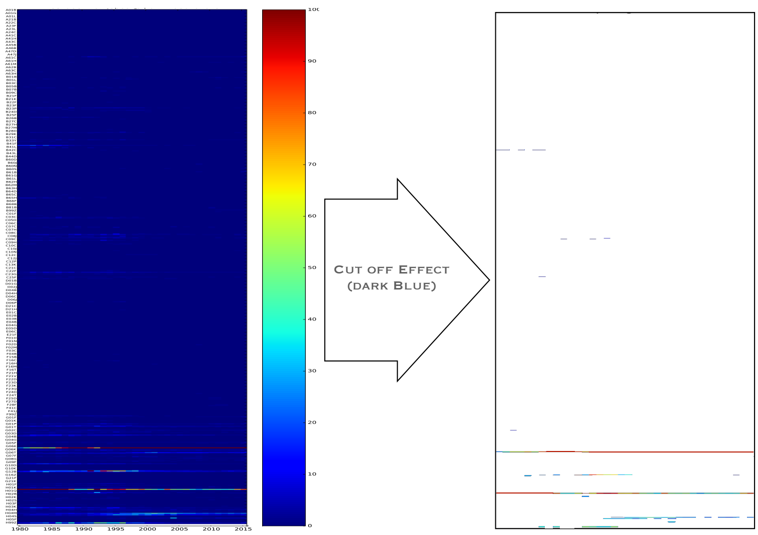

Figure 7.9: *TechSpectrogram* of IBM corp. at *IPC-SubClass*, and its dark-blue Cutoff version (right-hand graph)

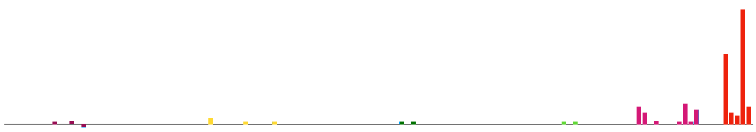

Figure 7.10: *TechSpectrum* of Apple minus the *TechSpectrum* of Microsoft both at *IPC-Class* level.

At the level of *technology spectrum*, its potential for providing information, easy to understand, about a technical field or a tech company in relation to another field or company seems promising. Let's see some cases of elementary processing of two spectrums to perceive this potential for providing additional information. The idea of these cases is to highlight the differences between the development activities of two companies of the same sector. The first case is using the spectrum of Apple and Microsoft, and the second case is the spectrum of Toyota and Ford Motors, two reference companies in computing and vehicle technology respectively.

In the first case, the *technology spectrums* of Apple and Microsoft will be compared at *IPC Class* level. In fact, we are just going to subtract bin-by-bin the spectrum of Microsoft from the spectrum of Apple to highlight their differences. We note by doing so that some bins can result in a positive or a negative value depending on whether Apple has more development activities than Microsoft or the other way round. Observing the result of this bin-by-bin subtraction of spectrums (See Figure 7.10), it seems clear at first look that the main difference between both companies is in the red area [6], in particular in *H04* and *H01*, corresponding to electric communication techniques and basic electric elements respectively. If now, to reveal in detail which techniques are behind these generic classes, we do the same exercise at a more refined level of resolution, the *IPC-SubClass* level, the result is that the major differences between both companies appear in H04W and H01L, corresponding to wireless communication networks and semiconductor devices respectively.

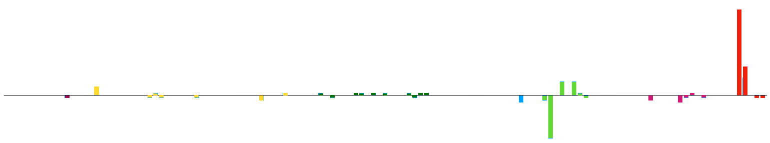

Figure 7.11: *TechSpectrum* of Toyota Motors minus the *TechSpectrum* of Ford Motors both at *IPC-Class* level.

In the second case, the *technology spectrums* of *Toyota Motors* and *Ford Motors* will also be compared at *IPC Class* level. As in the previous case, the spectrum of Ford Motors will be just subtracted from the spectrum of Toyota Motors. In Figure 7.11 We show a graph with the result of this subtraction of spectrums per bin. As in the previous example, the subtraction can result in bins with positive values, indicating more development activities of *Toyota* than *Ford Motors*, or it can result in bins with negative values, indicating more development activities of *Ford Motors* than *Toyota*. This graph shows significant differences between Toyota and Ford in two components, *H01*, with positive value and *F02*, with negative value. These bins correspond to electric elements and combustion engines respectively. Going deeper into the classification hierarchy to have more detail information, we can see that the H01 difference is due to a significant larger number of developments of Toyota in *H01M* – battery technology – and in a lesser extend *H01L* – semiconductors –. On the other hand, *Ford Motors* has a significant larger number of developments in *F02M* and *F02B*, corresponding to carburettors and combustion engines respectively.

[6]The red area of the spectrum is the “H” *IPC-Section* corresponding to electricity

The previous processing examples of *technology spectrums* and *spectrograms* only outline their potential for characterising the *piece of technology* represented by our two new tools. Additionally, in this PhD Research, we have developed two original applications using our new data structures and graphical tools.

The first application discloses a new method to assess the integration level of a technical field (See published paper **P3** or chapter 6 of this thesis). This method uses the *technology footprint* and the *technology spectrum* in combination with citation networks analysis, in particular *CitNets*. Three case studies are disclosed, namely *computer generated animation*, *regenerative medicine* and *computer tomography*. In our procedure, the most dominant technological components of the technical field are identified using its *technology spectrum*. Then *CitNets* analysis is used to identify groups of patents that are related within the technical fields based on citation relationships. In *CitNets* groups of patents that reference the same prior art tend to be clustered together, such clusters represent subtopics within the technical fields. The basic idea behind our method is to evaluate the existence of non-connected subtopics in different components. Our procedure combines both tools, the *spectrum* and *CitNets*, in a series of back and forth steps between them. In each loop, the level of intertwining of the main technological components is evaluated and an indication of *mono–*, *multi–* or *inter–disciplinarity* nature of the technology is obtained at the end of the procedure.

In the field of the history of technology, our procedure can facilitate the study of the evolution of a specific technology. The monitoring of the changes in the technological components inter-correlations can be a helpful tool to analyse the dynamics of the maturing of a technology. Our procedure in the hands of technology managers could accelerate the understanding of a technology under study by providing an assessment of its *mono–*, *multi–* or *inter–disciplinarity* nature, and therefore a tool providing such information could be beneficial to identify the skills required in a research teams. Such a tool could also help to perceive the kind of knowledge that a specific technology gathers within a company or research center under analysis, for instance, in the case of the analysis of the technologies owned by a company in view of merge and acquisition actions.

The second application generates technology maps based on technology proximity computed by processing our new data structures (See under submission paper in Appendix **A - P4** of this thesis). Two new graphs are disclosed, the *technology maps* and the *focused technology maps*. We have illustrated our new maps with to technical fields, namely Medical and automotive technology.

The first graph, the technology map – *TechMap* –, visualises a set of given technologies, technical fields and tech companies positioning them in relation to each other. The distance between items to visualise, in our case these are *pieces* *of technology*, is computed with the *soft cosine* algorithm, wherein the features are the IPC codes. In consequence, each piece of technology to be visualised is represented by a vector where the coordinate values correspond to each IPC-code's amounts of its technology spectrum. An algorithm of Multi-Dimensional Scaling – *MDS* – is employed to generate the coordinates of each item according, as much as possible, to the distance matrix. MDS visualises the items in a 2D plane based on the distance information using a nonlinear dimensionality reduction, the distances between the elements are the key information to gather.

Our second graph, the *Focused Technology Map* generates a visualisation of the given technologies in relation to a selected one, the *focus*, which is located at the centre of a circle with drawn radii corresponding to every IPC code. In this graph, the items are positioned at the computed distances from the focus, and on the radius representing its IPC code with the highest figures.

7.5 Relationships and Findings of the Compendium Papers

The present PhD research started by identifying and analysing the special role played by two bodies of patent literature, namely the prior art patent citations and the classification codes assigned to patent publications. The identification of this role is at the origins of this PhD thesis research, and led to the definition and developments of new tools for exploiting this role to characterise technology (see Introduction chapter).

Figure 7.12 is a scheme of the relation between the four papers forming this PhD thesis by compendium, the three published papers (**P1**, **P2** and **P3**) and the fourth paper under submission (See Appendix **A - P4**). As can be seen, the research and the corresponding paper follow a certain sequential order.

The first published paper, **P1**, studied how a particular set of patent publications which represents for a given period of time a *piece of technology*, namely a technical field, a tech company or a researcher (or group of researchers), can be characterised or described by the classification assigned to the particular set of patents and its prior art patent citations. As a consequence of this analysis, the role as a footprint of technology of prior art patent citations in combination with patent classification was identified, and a new data structure – the *technology footprint* – and its correspondent graphical tool – *technology spectrum* – were defined. Two case studies are developed in this paper, the first case discloses the inventor of the computer "mouse", Mr. Douglas Engelbart. The second case focuses in the early year of computerised tomography.

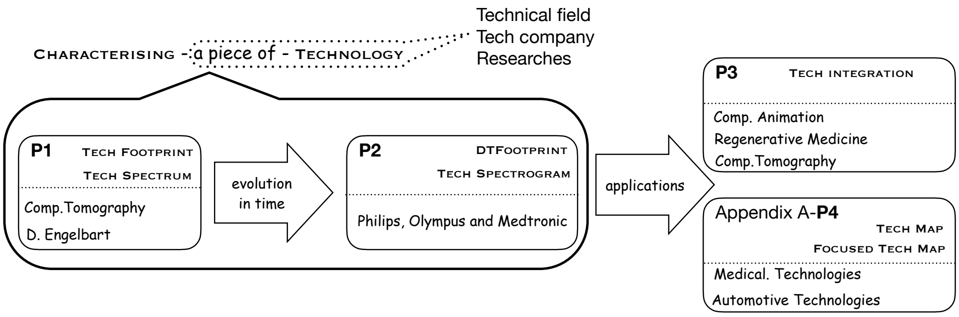

Figure 7.12: Overview of the interrelation between the papers forming this PhD thesis by compendium

In order to complement the analysis made in **P1** that covers a complete period of time, we have explore the evolution in time. So, the second published paper, **P2**, departing from the analysis made in **P1**, and highlighting the structural similarity to spectral representation of sound waves, explores the potential to analyse and visualise the evolution in time of a particular *piece of technology.* As a consequence of this analysis, another new data structure – the *Dynamic Technology Footprint* – and its correspondent graphical tool – the *technology spectrogram* – are defined. Both, the data structure and the graphical tool, are focusing on the evolution of the technology under study. In this paper some case studies are also explored. The evolution of a medical technical field, in our case *computerised tomography – C.T.* –, and three companies also related to the medical field, *Koninklijke Philips* N.V., *Medtronic* corporation and *Olympus* corporation. The time interval of observation is in all our cases goes from the years 1980 to 2015 (both included).

The third published paper, **P3** and the paper under submission, in Appendix **A - P4**, disclose two different applications of the results from the first two published papers, their two new data structures and the correspondent graphical tools.

P3 outlines a new method to assess whether a technology defined by a set of patents fits into one of the three integrative concepts, analogous to those applied to academic research, namely mono-, multi- or inter-disciplinarity. This method uses the *technology spectrum* in combination with citation network analysis, in particular the *CitNets* tool. The outlined method is illustrated with three technical fields, computer tomography, regenerative medicine and computer-generated animation.

Furthermore, Appendix **A - P4** presents the use of our two new data structures for exploring technology proximity. Two new graphical tools using our new data structures are defined, the *Technology Map* and the *Focused Technology Map.* Two case studies are developed, medical technologies and automotive technologies.

7.6 Papers' Specificities, Contributions and Results

7.6.1 Published paper P1

The first published paper, **P1**, develope two case studies to illustrate the characterisation of a technology based on patent information. The first, the technology developed by the inventor of the mouse, Mr. Douglas C. Engelbart, and the second, the three-dimensional reconstruction from X-ray projections, a technology known as *computer tomography.* In the first case, the technology is represented by the set of patents in which Mr. Engelbart appears as the inventor or one of the inventors. In the second study case, the technology is represented by the set of patents assigned to the IPC code *A61B6/03*.

In the two cases presented in this publication, the data gathering follows the data collection path (see section "*Data collection*" in chapter 3 - *Methodology*). The database used to collect the patents is *ESPACENET*. In the first case, there is not a time limit imposed to the patents to be collected because the constraint is the existence of patents with Mr. Engelbart in the inventor's field. On the contrary, the second case, the *computerised tomography* has a time limit constraint to the first collection of having a priority date field from 1970 to 1979 (both included) in a series of two years' time. For example, the first set of patents is constrained to have its priority date in 1970 or 1971, the second set is constrained to have its priority date in 1972 or 1973, and so on. This time constraint is implemented because the aim of this case is to study changes in the early years of the computer tomography technology. On top of this time limitation, we select only patents classified in the IPC code *A61B6/03* to gather only patents pertaining to the computer tomography field.

The contributions of **P1** are the identification and analysis of the role of patent citations and classification as a footprint of technology, and the definition of a new data structure, the technology footprint, and its corresponding graphical tool the technology spectrum. Additionally, it is identified the role played by the different values of the technology spectrum as technological components of the technology under study. The first case, Mr Engelbart's case, has illustrated the usefulness of our new tools to characterise the technology developed by an inventor by revealing the range of his developments beyond the invention for what he is universally known, the computer mouse. The second case, the *CT* case, has shown how our new tools can be used to characterise the origins of a given technology. Both cases have revealed the interest of its use in fields such as the history of technology and technology management.

7.6.2 Published paper P2

The second published paper, **P2**, departing from the new data structure defined in **P1**, underline the formal similarities of our tool with the frequency spectral representation of sound waves replacing the *frequency components* of the sound wave by the *technological components* of the technology under study. Consequently, a second data structure, the *DTFootprint*, and the corresponding graphical tool, the *technology spectrogram* are defined for analysing and visualising the evolution in time of the *technological components* of a specific technology. The time basis for this spectrogram is defined to 1 year.

At an experimental level we have analysed the same medical technology that in **P1**, computerised tomography. We have also analysed three companies with important development activities in the medical field, namely *Koninklijke Philips N.V.*, *Medtronic corporation* and *Olympus corporation*, as examples to explore the new data structure and graphical tool. These experimental cases are studied along the time interval from 1980 to 2015.

In all the case studies presented in **P2** the data gathering follows the common data collection path (see section "*Data collection*" in chapter 3 - *Methodology*). The database used for the collection of patents is in these cases *PATSTAT*. The first collection of patents is constrained to have the priority date field from 1980 to 2015 (both included), and they are gathered on an annual basis. In other words, the *DTFootprint* (or the graphical tool, the *technology spectrum*) is formed by gathering annual *technology footprints* corresponding to the technology under study and sequencing them from 1980 to 2015. The generation of the Technology spectrogram graphs are implemented in python using the library *Matplotlib*.

The contribution of **P2** is the definition of a new data structure, the *Dynamic Technology Footprint*, and its correspondent graphical tool, the *technology spectrogram*. These two entities are the dynamic version of the ones defined in **P1**, the *technology footprint* and the *technology spectrum*. The application of our new data structure and graphical tool to the history of technology and technology management has been explored throughout our practical cases of study.

7.6.3 Published paper P3

The third published paper illustrates the role and use of the new graphical tool, the *technology spectrum* in assessing technology integration. This publication outlines a new method to evaluate whether a technology fits into one of the three integrative concepts, analogous to those applied to academic research, namely mono—, multi— or inter—disciplinary. The method introduced in this publication uses our new *technology spectrum* in combination with citation network analysis.

The publication developed three archetypical case studies: computer-generated animation, regenerative medicine and computer tomography.

The data gathering for building the *technology spectrum* follows the common data collection path (see section "*Data collection*" in chapter 3 - *Methodology*). The database used for collection the patents in these cases is Derwent WPI[7] , and the citation network analysis is done through *CitNets*.

This third publication contributes by illustrating the use of our new tools but mainly outlining a new method for technology integration for assessing the mono-, multi- or inter-disciplinarity nature of a given technology. This method can help to evaluate the maturity of a technology.

7.6.4 Under Submition paper P4

This thesis also comprises a paper "under submission". The text, hereby incorporated in the Appendix **A - P4**, illustrates the role and use of our new data structures and tools for computing technology proximity between given technologies.

At experimental level, some tech companies are studied in **P4**. A first group of medical tech companies formed by *Boston Scientific*, *GE–medical*, *Medtronic*, *Nihon Kohden*, *Olympus medical*, *Omrom medical* and *Siemens HealthCare*. A second group with companies from computer science and telecom composed by *Apple*, *Cisco*, *Ericsson*, *IBM* and *Microsoft*. A final group with companies from vehicle technology constituted by *Ford motors*, *General Motors*, *Hyundai*, *Nissan*, *Toyota motors* and *Volkswagen*.

In all the experimental cases presented the data gathering follows the common above-mentioned data collection path (see section "*Data collection*" in chapter 3 - *Methodology*). The database used to collect the patents is in these cases *PATSTAT*. The *first collection* of patents are constrained to have the priority date field from 1980 to 2015 (both included). A matrix of distances between every and each

[7]DWPI as Derwent World Patent Information

item to be visualised in the *TechMap* is computed using the *cosine distances* of the respective *technology footprints*. The dimensionality reduction is implemented using *MultiDimensional Scaling – MDS –*, and a 2D graph is generated visualising the items as much as possible at its respective relative distances. The *MDS* function is implemented using the python library *SKlearn*.

The fourth paper, **P4** contributes by defining a technology proximity computation based on our two new data structures. This paper also contributes by defining two new graphical tools, the *Technology Map* and the *Focussed Technology Map* for visualising technology relationships. The *Technology Map* visualises technology items (technical fields, tech companies or groups of researchers) in a 2D plot in which the items appear positioned relative to each other. The *Focussed Technology Map* is a radial graph which visualises technology items with reference to a selected one, called the focus, which is located at the centre of a circle whose radii correspond to the complete set of IPC codes. In this way, the items provide information by its angular and radial location.

7.7 Summary

Overall, the research questions and objectives of this Thesis (See chapter 1 - *Introduction*, section 1.3 *Research questions and objectives*) were fulfilled. Throughout the papers (**P1** to **P4**) and its discussion presented in this Thesis by compendium , we have outlined the characterisation of a *piece of technology*, namely by finding its technological foundations, its evolution, and evaluating the inter-relations of these technologies with any other existing technology. This *piece of technology* was defined as the set of patents assigned to a particular technical field, owned by a tech company, or invented by a group of researchers. The identification of the technological components of a *piece of technology* was disclosed and some case of study were presented, as well as the disclosure of its evolution and inter-relation.

Chapter 8

Conclusions and future works

8.1 Conclusions

In this thesis, we set out to understand whether the processing of information from a set of patents defining a piece of technology can characterise it. We wanted to discover (1) what were the technological components of a piece of technology? (2) how the technological components of a technology evolve in time. (3) How are the technological components inter–related?, and (4) What is the technological proximity of a given piece of technology to other technologies.

We first identified three features of prior art patent citations that are unique within the body of bibliographic references in science and technology publications:

- first, *who* is the responsible for the list of prior art references , namely the patent examiner in place of the author,
- second, *why* they are selected to be cited, that is the evaluation of novelty and non-obviousness, and
- third, *where* they are classified within an universal classification scheme, the *IPC*, containing the whole spectrum of existing technologies.

The differences between prior art patent citations and bibliographic references in NPL have been analysed from a plurality of angles, such as the basis of the citations, the homogeneity of the browsed collection of documents, the citation features, and the structure of the cited documents. From this we found that a first set of differences, namely the patent unique identifier, a patent universal format for information fields, and the existence of a legal definition of patent authorship , facilitates the retrieval and processing of citations. A second set of identified differences , such as the specific reason to cite patent prior art, and the existence of a universal classification scheme, allows us to identify the role of prior art patent citations as a footprint of technology.

Furthermore, We have defined a *piece of technology* as the set of patents classified in a technical field, the patents owned by a tech company, the patents granted to an inventor, or any combination of the three previous possibilities. From our analysis of differences and and singularities of patents and prior art citations we have identified the special role that patent information, contained in that set of patents and its corresponding prior art citations, can play for characterizing that *piece of technology*

We subsequently have defined two new data structures and its corresponding graphical tools that exploit the role of these bodies as footprint of technology, and we use them to characterise, to describe, to better understand and to compare particular technical fields, tech companies or researchers. These new data structures and tools are inspired in the spectral characterization of sound waves, using classification codes in place of frequency functions. Therefore, we have named the data structures *Technology Footprint* and *Dynamic Technology Footprint*, and the corresponding graphical tools *Technology Spectrum* and *Technology Spectrogram* respectively.

The *Technology Footprint* is formed by selecting a set of patents (assigned to a technical field, owned by a tech company, or invented by an inventor or group of inventor) along a time interval, collecting its prior art, and forming a final collection with both sets, the selected set of patents and its prior art citations. Then, the classification codes allocated to this final collection are grouped and ordered according to the *IPC*. The second data structure, the *Dynamic Technology Footprint*, results from the sequencing of yearly *Technology Footprints* along a number of years. These data structures and graphical tools provide us information about the *technology components* of the *piece of technology* under study, and allows to analyse its evolution.

In order to solidify our constructs and tools, we undertook two case studies to illustrate the *Technology Footprint* and the *Technology Spectrum*: The inventor of the computer mouse, Mr. Douglas C. Engelbart and the first years of *Computer Tomography* technology. In turn, the following case studies were presented to show the formation of the *Dynamic Technology Footprint* and the *Technology Spectrogram*: Three tech companies, *Philips*, *Olympus* and *Medtronic* and again *C.T.* technology.

Finally, two application of our new data structures and tools were developed to explore their use and potential. The first application is a new method for assessing technology integration that uses our new *Technology footprint* together with citation networks analysis – *CitNets* –. We have introduced a new algorithm to asses whether a given technology fits into one of the integrative concepts, analogous to those applied in academic research, namely mono-, multi- and inter-disciplinarity. We have tested the potential applicability of our procedure with three case studies. The scientific contribution of our procedure is to provide a methodological framework and the corresponding indicators to decision-makers and historian of technology to asses how interdisciplinary a technology is at a certain interval of time. The second application is the generation of technology maps – *TechMaps* – that applies our new *Technology footprint* and *Dynamic Technology footprint* to explore technology proximity. Our new application defines two new complementary maps. We have shown in several case studies how our tools facilitate the study of the evolution of specific technical fields, and the tracing of the divergence or convergence of tech companies. The relative positioning of technical fields and tech companies helps to better identify those that have some techniques in common, as well as to improve the understanding of the technical developments and trends in firms, as we have illustrated for motor car companies moving from pure mechanics to electrical technologies.

Both the presented case studies and the developed applications disclose the procedure to generate the new data structures and tools from the database querying, as well as the usefulness of our new tools in history of technology research, and technology management. We would like to note that this PhD research is constrained to applied technology originated mainly by industry due to the fact that the source of information are patent publications. On the other hand, it is important to highlight that the present research does not cover the chemical technology fields. The extension to these fields is foreseen for future works.

As a result of this PhD research, the new data structures, its corresponding graphical tools and the disclosed cases of study and applications a step towards characterising technology using patent information, our main goal, is made. The definition of our new data structures and its corresponding graphical tools will help historian and managers of technology to better and easily understand technical fields and tech companies.

8.2 Future Works

At present we are trying to model the dynamics of the *technology components* in order to study the emergence of new technologies or technological events. We are looking to model the "forces" behind the temporary changes.

With respect to the *DTFootprint* and the corresponding *technology spectrograms* we are trying to identify patterns such as, among others, increasing and decreasing *ramps*, *peaks*, *bell* shapes, *plateau* or *step - up and - down* shapes. The identification of these patterns could highlight some specific situations, such as a fronts of development or a loss of interest for a technical area in a particular moment or the appropriation through some acquisition of a specific know-how. Our idea is to identify all these shapes , catalogue them , and look for potential correlations between the patterns and technological and business events.

We are also studying image processing developed for frequency spectrograms analysis and its translation to our *technology spectrograms* in order to identify common patterns and to analyse their meaning.

The method for assessing technology integration outlined in **P3** (and the corresponding chapter 6) will be validated with an extended set of technologies to evaluate the span of technical fields of our indicator and to fine tune the two used thresholds. We are also testing our procedure with a shorter time interval to study how and when two different disciplines fuse to improve the understanding of the evolution and dynamics of the maturity process of a technology.

Finally, we are testing alternative distance computation, and the improvement of the similarity matrix by using text based similarities and citation network analysis. We will experiment with some algorithms for the automatic clustering of items within *Technology Maps*. Moreover, we foresee the use of predictive modelling of the technology maps dynamics to anticipate the evolution of the relative position of tech companies and technical fields.

Further research will be oriented to compute technology maps with smaller time intervals to improve the perception of the maps evolution and to investigate the modelling of trends in technology fields and tech companies.

Chapter 9

Bibliography

9.1 Bibliography

Albert, J. (2016). *Measuring Technology Maturity – Operationalizing Information from Patents, Scientific Publications, and the Web*. Springer Gabler. doi:`https://doi.org/10.1007/978-3-658-12132-7`.

Alstott, J., Triulzi, G., Yan, B., and Luo, J. (2017). Mapping technology space by normalizing patent networks. *Scientometrics*, 110:443–479.

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

Azizian, N., Sarkani, S., and Mazzuchi, T. (2009). A comprehensive review and analysis of maturity assessment approaches for improved decision support to achieve efficient defense acquisition. *Proceedings of the World Congress on Engineering and Computer Science 2009*, II.

Basberg, L. (1987). Patents and the measurement of technological change: A survey of literature. *Research Policy*, 16:131–141.

Blackwell, A., Wilson, L., Street, A., Boulton, C., and Knell, J. (2009). Radical innovation: crossing knowledge boundaries with interdisciplinary teams. *Technical Report. University of Cambridge, Computer Laboratory*, Number–760.

Borg, I. and Groenen, P. (2005). *Modern multidimensional scaling: Theory and applications.* Springer.

Borner, K. and Scharnhorst, A. (2009). Visual conceptualizations and models of science. *Journal of Informetrics*, 3:161–172.

Bowen, Y. and Luo, J. (2016). Measuring technological distance for patent mapping. *Journal of the Association for Information Science and Technology*, 68:423–437.

Boyack, K., Wylie, B., and Davidson, G. (2002). Domain visualization using vxinsight® for science and technology management. *Journal of the American Society for Information Science and Technology (JASIST)*, 53(9):764–774.

Boyack, K., Wylie, B., Davidson, G., and Johnson, D. (2000). Analysis of patent databases using vxinsight. *In 9th International Conference on Information and Knowledge Management*.

Camacho-Minano, M. and Nunez-Nickel, M. (2009). The multilayered nature of reference selection. *Journal of the Association for Information Science and Technology*, 60(4):754–777.

Campbell, L. (2005). Overcoming obstacles to interdisciplinary research. *Conservation Biology*, 19(2):574–577. doi:`https://doi.org/10.1111/j.1523-1739.2005.00058.x`.

Carpenter, M. and Narin, F. (1983). Validation study. patent citations as indicators of science and foreign dependence. *World Patent Information*, 5(3):180–185. doi:`https://doi.org/10.1016/0172-2190(83)90139-4`.

Caspersen, K., Madsen, M., Eriksen, A., and Thiesson, B. (2017). A hierarchical tree distance measure for classification. (pp. 502-509). *In 6th International Conference on Pattern Recognition Applications and Methods*, pages 502–509.

Chang, P. (2010). Using patent analysis to monitor the technological trendds in an emerging field of technology: a case of carbon nanotube field emisision display. *Scientometric*, 82:5–19.

Chen, S., Huang, M., and Chen, D. (2011). Visualization of the technology evolution in smart grid. *Proceedings of PICMET 2011*, 5:1–7.

Chubin, D. and Moitra, S. (1975). Content analysis of references: Adjunct or alternative to citation counting? *Social Studies of Science*, 5(4):423–441.

Colange, B., Vuillon, L., Lespinats, S., and Dutykh, D. (2019). Interpreting distortions in dimensionality reduction by superimposing neighbourhood graphs. *Proceedings of IEEE Visualization Conference*, pages 211–215.

Cronin, B. (1984). The citation process: The role and significance of citations in scientific communication. *The Citation Process*, ISBN 0 947568 01 8.

de Rassenfosse, G., Dernis, H., and Boedt, G. (2014). An introduction to the patstat database with example queries. *ArXiv*, abs/1404.7447.

Ellis, P., Hepburn, G., and Oppenhein, C. (1978). Studies on patent citation networks. *Journal of Documentation*, 34(1):12–20.

Ellis, P., Hepburn, G., and Oppenhein, C. (2000). Studies on patent citation networks. *Journal of Documentation*, 34(1):12–20.

EPO (2016). Invertors' handbook. *www.epo.org/learning-events/materials/inventors-handbook.html*.

EPO (2019). Patstat. *At: www.epo.org/searching-for-patents/business/patstat.html*.

EPO (2022a). Cooperative patent classification. *https://www.epo.org/searching-for-patents/helpful-resources/first-time-here/classification/cpc.html*.

EPO (2022b). What is a patent. *e-courses.epo.org*.

EPO (2022c). What is patent information? *https://www.epo.org/searching-for-patents/helpful-resources/first-time-here.html*.

Erdi, P., Makovi, K., Somogyvári, Z., Strandburg, K., Tobochnik, J., Volf, P., and Zalanyi, L. (2013). Prediction of emerging technologies based on analysis of the us patent citation network. *Scientometrics*, 95(1):225–242. doi:`https://doi.org/10.1007/s11192-012-0796-4`.

Ernst, H. (2003). Patent information for strategic technology management. *World Patent Information*, 25:233–242.

Freeman, C. and Soete, L. (2009). Developing science, technology and innovation indicators: What we can learn from the past. *Research policy*, 38(4):583–589.

Garfield, E. (1955). Citation indexes for science: A new dimension in documentation through association of ideas. *Science*, 122(3159):108–111.

Garfield, E. (1956). Citation indexes: new paths to scientific knowledge. *The Chemical Bulletin*, 43(4):108–111.

Garfield, E. (1957). Breaking the subject index barrier - a citation index for chemical patens. *Journal of the Patent Office Society*, 39(8):583–595.

Garfield, E. (1962). Can citations indexing be automated? *Essay of an Information Scientisty*, 1:84–90.

Geisler, E. (2000). *The Metrics of Science and Technology*. Quorum Books, Greenwood Publishing Group Inc.

Godin, B. (2004). The obsession for competitiveness and its impact on statistics:the construction of high-technology indicators. *Journal of the Patent Office Society*, 33(8):1217–1229. doi:https://doi.org/10.1016/j.respol.2004.07.005.

Guarino, N. (1997). Understanding, building and using ontologies. *International Journal of Human Computer Studies*, 46(2):293–310. doi:https://doi.org/10.1006/ijhc.1996.0091.

Haupt, R., Kloyer, M., and Lange, M. (2007). Patent indicators for the technology life cycle development. *Research Policy*, 36(3):387–398. doi:https://doi.org/10.1016/j.respol.2006.12.004.

Hepp, M. (2007). Possible ontologies: How reality constrains the development of relevant ontologies. *IEEE Internet Computing*, 11(1):90–96. doi:10.1109/MIC.2007.20.

Huutoniemi, K., Klein, J., Bruun, H., and Hukkinen, J. (2010). Analyzing interdisciplinarity: Typology and indicators. *Research Policy*, 39:79–88. doi:https://doi.org/10.1016/j.respol.2009.09.011.

Igami, M. (2008). Exploration of the evolution of nanotechnology via mapping of patent applications. *Scientometrics*, 77(2):289–308.

Israel, P. and Rosenberg, R. (1992). Patent office records as a historical source: The case of thomas edison. *Technology and Culture*, 32:1094–1101.

Jaffe, A. (1986). Technological opportunity and spillovers of r and d: evidence from firms patents, profits and market value. *American Economic Review*.

Jaffe, A. and Trajtenberg, M. (1999). International knowledge flows: Evidence from patent citations. *Scientometrics*, 8(1-2):105–136.

Jang, S. (2009). How do latecomers catch up with forerunners? analysis of patents and patent citations in the field of flat panel display technologies. *Scientometrics*, 79(3):563–591.

Joo, S. (2010). Measuring relatedness between technological fields. *Scientometrics*, 83(2. doi: 10.1007/s11192-009-0108-9):435–454.

JPO (2022). Fi–fterm list. *https://www.jpo.go.jp/e/system/patent/gaiyo/seido-bunrui/index.html*.

Khattak, A., Latif, K., Lee, S., and Y. (2009). Ontology evolution: A survey and future challenges. *Communications in Computer and Information Science*, 62:68–75. doi:https://doi.org/10.1007/978-3-642-10580-7_11.

Khattak, A., Pervez, Z., Lee, S., and Y. (2010). After effects of ontology evolution. *Proceedings of the 5th International Conference on Future Information Technology*, pages 1–6. doi:10.1109/FUTURETECH.2010.5482765.

Kothari, C. (2004). *Research methodology methods and techniques.* New Age International Publishers.

Kwasnik, B. (1992). "The role of classification structures in reflecting and building theory". *Advances in Classification Research Online*, pages 63–82. Accessed: 2018-08-20.

Lee, C., Cho, Y., Seol, H., and Park (2011). A stochastic patent citation analysis approach to assessing future technological impacts. *Technological Forecast and Social Change*, 79(1. doi: 10.1016/j.techfore.2011.06.009):16–29.

Lespinats, S., Verleysen, M., Giron, A., and Fertil, B. (2007). Dd-hds: A method for visualiza-tion and exploration of high-dimensional data. *IEEE Transactions on Neural Networks*, 18(5):1265–1279.

Leydesdorff, L. (2008). Patent classifications as indicators of intellectual organization. *Journal of the American Society for Information Science and Technology*, 59(10):1582–1597. doi:https://doi.org/10.1002/asi.20814.

Leydesdorff, L. (2018). Diversity and interdisciplinarity: how can one distinguish and recombine disparity, variety, and balance? *Scientometrics*, 116:2113–2121.

Leydesdorff, L., Alkemade, F., Heimeriks, G., and Hoekstra, R. (2015). Patents as instruments for exploring innovation dynamics: geographic and technological perspectives on photovoltaic cells. *Scientometrics*, 102(1. doi:10.1007/s11192-014-1447-8):629–651.

Leydesdorff, L., Carley, S., and Rafols, I. (2013). Global maps of science based on the new web-of-science categories. *Scientometrics*, 94:589–593.

Leydesdorff, L., Duncan, K., and Ismael, R. (2008). Interactive overlay maps for us patent (uspto) data based on international patent classification (ipc). *Scientometrics*, 98(3):1583–1599. doi:https://doi.org/10.1007/s11192-012-0923-2.

Lezama–Nicolás, R., Rodríguez–Salvador, M., Río–Belver, R., and Bildosola, I. (2018). A bibliometric method for assessing technological maturity: the case of additive manufacturing. *Scientometrics*, 117:1425–1452. doi:https://doi.org/10.1007/s11192-018-2941-1.

Li, Y., Bandar, Z., and McLean, D. (2003). An approach for measuring semantic similarity between words using multiple information sources. *IEEE Transactions* *on Knowledge and Data Engineering*, 15(4):871–882. doi:`https://doi.org/10.1109/TKDE.2003.1209005`.

Liu, W. and Liao, H. (2017). A bibliometric analysis of fuzzy decision research during 1970–2015. *International Journal of Fuzzy Systems*, 19(1):1–14.

Liu, Z. and Zhu, D. (2009). Web mining based patent analysis and citation visualization. *Proceedings*, pages 19–23.

Ma, Y., Gu, X., and Wang, Y. (2010). Histogram similarity measure using variable bin size distance. *Computer Vision and Image Understanding*, 114:981–989.

MacRoberts, M. and MAcRoberts, B. (1989). Problems of citation analysis: A critical review. *Jasist*, 40(5):342–349.

Martinelli, A. (2012). An emerging paradigm or just another trajectory? understanding the nature of technological change using engineering heuristic in the telecommunications switching industry. *Applied Sciences*, 41(2):414–429.

Martinelli, A. and Nomalerr, O. (2014). Measuring knowledge persistence: a genetic approachto patent citation networks. *J Evol Econ*, 24:623–652.

Matsuura, J. (2008). *Jefferson vs. the Patent Trolls.* University of Virginia press.

Mejia, C. and Kajikawa, Y. (2017). Bibliometric analysis of social robotics research: Identifying research trends and knowledgebase. *Research Policy*, 7:1–17.

Merriam-Webster (2022). Merriam-webster online: Dictionary and thesaurus. *at http://http://www.merriam-webster.com/*. Last consultation 10-02-2022.

Millar, V. (2016). Interdisciplinary curriculum reform in the changing university. *Teaching in Higher Education*, 21(4):471–483. doi:`https://doi.org/10.1080/13562517.2016.1155549`.

Moed, H. F., Glänzel, W., and Schmoch, U. (2004). Handbook of quantitative science and technology research. In *The Use of Publication and Patent Statistics in Studies of S&T Systems*. Springer.

Mogee, M. and Kolar, R. (1994). International patent analysis as a tool for corporate technology analysis and planning. *Technol. Anal. Strateg. Manag*, 6(4):485–504.

Morillo, F., Bordons, M., and Gomez, I. (2003). Interdisciplinarity in science: A tentative typology of disciplines and research areas. *Journal of the American Society for Information Science and Technology*, 54(13):1237–1249. doi:`https://doi.org/10.1002/asi.10326`.

Morris, S., DeYongb, C., Wua, Z., Salmanb, S., and Yemenu, D. (2002). Diva: a visualization system for exploring document databases for technology forecasting. *Computers and Industrial Engineering*, 1(43):841–862.

Morris, S. and Yen, G. (2004). Crossmaps: Visualization of overlapping relationships in collections of journal papers. *Proceedings of the National Academy of Sciences of the United States of America*, 101:5291–5296.

Mossoff, A. (2007). Who cares what thomas jefferson thought about patents - reevaluating the patent privilege in historical context. *Cornell Law Review*, 92.

Murugesan, P. and Moravcsik, M. (1978). Variation of the nature of citation measures with journals and scientific specialties. *Jasist*, 29(3):141–147.

Narin, F. (1995). Patents as indicators for the evaluation of industrial research output. *Scientometrics*, 34(3):489–496. doi:`https://doi.org/10.1007/BF02018015`.

Nemet, G. and Johnson, E. (2012). Do important inventions benefit from knowledge originating in other technological domains. *Research Policy*, 41(1):190–200.

Nicolson, C., Starfield, A., Kofinas, G., and Kruse, J. (2002). Ten heuristics for interdisciplinary modeling projects. *Ecosystems*, 5:376–384. doi:`https://doi.org/10.1007/s10021-001-0081-5`.

Nolte, W. (2008). *Did I ever tell you about the whale? Or measuring technology maturity*. Information Age Publishing.

Oikawa, K. (2017). Inter-firm technological proximity and knowledge spillovers. *Tokyo Center for Economic Research*, Paper No. E-114.

Okubo, Y. (1997). Bibliometric indicators and analysis of research systems: Methods and examples. *OECD Science, Technology and Industry Working Papers*, 1997/01.

Osborne, P. (2015). Problematizing disciplinarity, transdisciplinary problematics. *Theory, Culture and Society*, 32(5–6):3–35. doi:`https://doi.org/10.1177/0263276415592245`.

Pereira, C., Da-Silva, R., and Porto, G. (2015). The scientific information provided through patents and its limited use in scientific research at universities. *Brazilian Journal of Science and Technology*, 2(1):1–11. doi:10.1186/s40552-015-0007-y.

Perez-Molina, E. (2014). The technological roots of computer graphics. *Annals of the History of Computing*, 36(3):30–41. doi:10.1109/MAHC.2014.47.

Poteralska, B. (2002). Application of technology assessment for the needs of randd management. *Proceedings of the 12th European Conference on Innovation and Entrepreneurship*, pages 520–529.

Rafols, I. and Meyer, M. (2014). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, 82(2):263–287. doi:https://doi.org/10.1007/s11192-009-0041-y.

Rainey, M. (2014). Free sources for patent searching: A review. *Business information review*, 31(4):216–225. doi:https://doi.org/10.1177/0266382114562106.

Reingold, N. (1960). U.s. patent office records as sources for the history of inventions and technological property. *Technology and Culture*, 1:156–167.

Salatino, A., Thanapalasingam, T., Mannocci, A., Osborne, F., and Motta, E. (2018). The computer science ontology: A large-scale taxonomy of research areas. *Lecture Notes in Computer Science*, 11137:187–205. doi:https://doi.org/10.1007/978-3-030-00668-6_12.

Scherer, F. (1984). Using linked patent and sandd data to measure interindustry technology flows. *Chapter in Book: Rand, Patents, and Productivity*, pages 417–464.

Schmoch, U., Laville, F., Pianta, M., and Sirilli, G. (1994). The measurement of scientific and technological activities: Using patent data as science and technology indicators. *Patent Manual*.

Schoen, A., Villard, L., Laurens, P., Cointet, J., Heimeriks, G., and Alkemade, F. (2012). The network structure of technological developments; technological distance as a walk on the technology map. *Science and Technology Indicators Conference*.

Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3):425–465. doi:https://doi.org/10.1023/B:SCIE.0000018542.71314.38.

Seo, W., Kim, N., and Choi, S. (2016). Big data framework for analyzing patents to support strategic randd planning. *2nd Intl Conf on Big Data Intelligence and Computing*, pages 746–753.

Serratosa, F. and Sanfeliu, A. (2006). Signatures versus histograms: Definitions, distances and algorithms. *Pattern Recognition*, 39:921–934.

Sidorov, G., Gelbukh, A., Gómez-Adorno, H., and Pinto, D. (2014). Soft similarity and soft cosine measure: Similarity of features in vector space model. *Computación y Sistemas*, 18(3):491–504.

Simon, H. and Sick, N. (2016). Technological distance measures: new perspectives on nearby and far away. *Scientometrics*, 107:1299–1320.

S.L. Jang, Y. Y. and Wang, T. (2011). Emerging firms in an emerging field: an analysis of patent citations in electronic-paper display technology. *Scientometrics*, 89(doi: 10.1007/s11192-011-0448-0):259–272.

Smith, K. (2006). Measuring innovation. *in The Oxford Handbook of Innovation*. doi:https://doi.org/10.1093/oxfordhb/9780199286805.003.0006.

Smith, L. (1981). Citation analysis. *Library Trends*, 30:86–106.

Song, M., Heo, G., and Kim, S. (2014). Analyzing topic evolution in bioinformatics: investigation of dynamics of the field with conference data in dblp. *Scientometrics*, 101(DOI 10.1007/s11192-014-1246-2):397–428.

Spear, B. (2002). Virtual reality: patent review. *World Patent Information*, 24:103–109.

Steele, T. and Stier, J. (2000). The impact of interdisciplinary research in the environmental sciences: A forestry case study. *Journal of the American Society for Information Science*, 51(5):476–484. doi:https://doi.org/10.1002/(SICI)1097-4571(2000)51:5<476::AID-ASI8>3.0.CO;2-G.

Stojanovic, L., Maedche, A., Motik, B., and Stojanovic, N. (2002). User-driven ontology evolution management. *Lecture Notes in Computer Science*, 2473:285–300. doi:https://doi.org/10.1007/3-540-45810-7_27.

Strelkov, V. (2008). A new similarity measure for histogram comparison and its application in time series analysis. *Pattern Recognition Letters*, 29:1768–1774.

Sung, H., Yeh, H., Lin, J., and Chen, S. (2017). A visualization tool of patent topic evolution using a growing cell structure neural network. *Scientometrics*, 11(3):1267–1285.

Taebi, B., Correljé, A., Cuppen, E., Dignum, M., and Pesch, U. (2014). Responsible innovation as an endorsement of public values: the need for interdisciplinary research. *Journal of Responsible Innovation*, 1(1):118–124. doi:https://doi.org/10.1080/23299460.2014.882072.

Takano, Y., Kajikawa, Y., and Ando, M. (2017). Trends and typology of emerging antenna propagation technologies: Citation network analysis. *International Journal of Innovation and Technology Management*, 14(1):2872–2881.

Tenenbaum, J., de Silva, V., and Langford, J. (2000). A global geometric frameworkfor nonlinear dimensionalityreduction. *Science*, 290:2319–2323.

Tomov, D. and Mutafov, H. (1996). Comparative indicators of interdisciplinarity in modern science. *Scientometrics*, 37(2):267–278. doi:`https://doi.org/10.1007/BF02093624`.

Torgerson, W. (1952). Multidimensional scaling: Theory and method. *Psichometrika*, 17(4):401–419.

Trajtenberg, M. (1990). A penny for your quotes: patent citations and the value of innovations. *RAND Journal of Economics*, 21(1):172–187.

Tress, G., Tress, B., and Fry, G. (2005). Clarifying integrative research concepts in landscape ecology. *Landscape Ecology*, 20(4):479–493. doi:`https://doi.org/10.1007/s10980-004-3290-4`.

Urpa, L. and Anders, S. (2019). Focused multidimensional scaling: interactive visualization for exploration of high-dimensional data. *BMC bioinformatics*, 20(1):1–8. doi:`https://doi.org/10.1186/s12859-019-2780-y`.

VanDenBesselaar, P. and Heimeriks, G. (2001). Disciplinary, multidisciplinary, interdisciplinary – concepts and indicators -. *Proceedings 8th International Conference on Scientometrics and Informetrics - ISSI2001*, pages 705–716.

VanRijnsoevera, F. and Hessels, L. (2011). Factors associated with disciplinary and interdisciplinary research collaboration. *Research Policy*, 40:463–472. doi:`https://doi.org/10.1016/j.respol.2010.11.001`.

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. doi:`https://doi.org/10.1142/S0219525907000945`.

Vijvers, W. (1990). The international patent classification as a search tool. *World Patent Information*, 12(1):926–30.

Villanueva-Rivera, L., Pijanowski, B., Doucette, J., and Pekin, B. (2011). A primer of acoustic analysis for landscape ecologists. *Landscape ecology*, 26:1233–1246.

Wagner, C., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., Rafols, I., and Boerner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (idr): A review of the literature. *Journal of Informetrics*, 5(1):14–26. doi:`https://doi.org/10.1016/j.joi.2010.06.004`.

Wartburg, I. V., Teichert, T., and Rost, K. (2005). Inventive progress measured by multi-stage patent citation analysis. *Research Policy*, 34:1591–1607.

Weingart, P. and Padberg, B. (2014). *University Experiments in Inter–Disciplinarity*. Blelefeld Verlag.

Werman, M., s. Peleg, and Rosenfeld, A. (1985). A distance metric for multi-dimensional histograms. *Computer Vision, Graphics, and Image Processing*, 32:328–336.

WIPO (2003). Recommendation for the numbering of published patent documents. Retrieved February 02 2017 from: http://www.wipo.int/export/sites/www/standards/en/pdf/03-06-01.pdf.

WIPO (2022a). International patent classification guide. Retrieved June 06 2016.

WIPO (2022b). Minimum pct documentation. last consultation on 16-08-2022.

WIPO (2022c). Wipo guide to using patent information", wipo at. Accessed: 2022-30-07.

WIPO-2 (2022). "International patent classification". `http://www.wipo.int/classifications/ipc/en/`. Accessed: 2022-08-10.

Woerter, M. (2012). Technology proximity between firms and universities and technology transfer. *The Journal of Technology Transfer*, 37:828–866.

Wu, F., Li, R., Huang, L., Miao, H., and Li, X. (2017). Theme evolution analysis of electrochemical energy storage research based on citnetexplorer. *Scientometrics*, 110(DOI 10.1007/s11192-016-2164-2):113–139.

Wu, Z. and Palmer, M. (1994). Verb semantics and lexical selection. *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*, pages 133–138. doi:`https://doi.org/10.3115/981732.981751`.

Yan, B. and Luo, J. (2017). Measuring technological distance for patent mapping. *JASIST*, 68(2):423–437.

Yoon, B. and Magee, C. (2018). Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction. *Technological Forecasting and Social Change*, 132:105–117.

Yoon, B. and Park, Y. (2004). A text-mining-based patent network: Analytical tool for high-technology trend. *Journal of High Technology Management Research*, 15:37–50.

Yoon, J. and Kim, K. (2012). Trendperceptor: A property function based technology intelligence system for identifying technology trends from patents. *Expert Systems with Applications*, 39:2927–2938.

Young, F. (1987). *Multidimensional scaling. History, Theory and Applications.* LEA publishers.

Appendices

Appendix A

TechMaps: Exploring technology relationships through patent information based proximity (- P4 - under submission)

Eduardo Perez-Molina and Fernando Loizides.

TECHMAPS: EXPLORING TECHNOLOGY RELATIONSHIPS THROUGH PATENT INFORMATION BASED PROXIMITY

EDUARDO PEREZ MOLINA+ AND FERNANDO LOIZIDES*

May 30, 2022

Abstract. Our work provides a novel method for rich information discovery about the evolution of technical fields and company developments through patent relationships. A new exploratory method and graphical tool to discover technology proximity based on patent classification information are introduced. By technology we mean a technical field (defined by an International Patent Classification – IPC – code or a combination of them) or an organization (such as a tech company, research centre or institution). A single data structure is used for characterizing both technical fields and organizations, to visualize them as items of the very same body. This new method generates two graphs, a first graph, the *TechnologyMap* visualizes technology items in a 2D plot wherein technical fields and companies will appear positioned relative to each other. A second graph, the *Focused TechnologyMap* visualizes technology items with respect to a selected one, the *focus*, which is located in the centre of a circle whose radii correspond to the complete set of IPC codes. This article represents the process and algorithms used for production of the graphs, and solidifies the assumptions of validity by presenting two of the many successful test cases to which it was applied.

1. Introduction

Discovering the proximity between technical fields have been proved to be important for the global understanding of science and technology, facilitating for example knowledge technology transfer, research collaboration between institutions (Woerter, 2012) or the identification of technology opportunities(Jaffe, 1986). By discovering technology proximity we mean to assess the similarity of technologies, namely the share knowledge and techniques. In case of characterizing technology using patent information, proximity could be evaluated identifying the commonalities at classification level(Alstott et al., 2017;Jaffe, 1986; Simon and Sick, 2016). In other words, two technologies are closer to each other than a third one when the two first have more classification information in common in any way that the third one(Boyack, Wylie, Davidson, and Johnson, 2000; Okubo, 2017; Schoen et al., 2012; Woerter, 2012).

Revealing the technical fields where companies have industrial or research developments could play a key role in the decision making of technology players in areas such as analysis

+ Universidad Politécnica de Madrid (UPM).
* Cardiff University.

of competition(Schoen et al., 2012), firm collaboration(Simon and Sick, 2016) or mergers and acquisitions(Simon and Sick, 2016).

Our objective with the present work is the visualization of the relative position of technology, that is placing in a graph technical fields and tech companies with respect to each other. The idea is to use a single data structure based on patent classification information for characterizing both, technical fields and organizations (company, research institution or university lab), in order to visualize them as items of the very same body, namely technology. In this way, observations such as proximity or dynamics could be extended from technical fields and tech companies to each other.

We have characterized a technical field or an organization (company, research institution or university group) by a single data structure, the *TechSpectrum*, formed by the aggregation into bins[1] of the IPC codes assigned to a set of patents and its prior art mixed in a specific proportion, and ordered according to the IPC(Perez-Molina, 2018).

This binning and ordering is done at different IPC granularities generating a particular *TechSpectrum* for each IPC level[2] (see for example in Figure: 1 the *TechSpectrum* graph at IPC-SubClass level for propulsion of electric vehicles – IPC code *B60L* – and Toyota Motors).

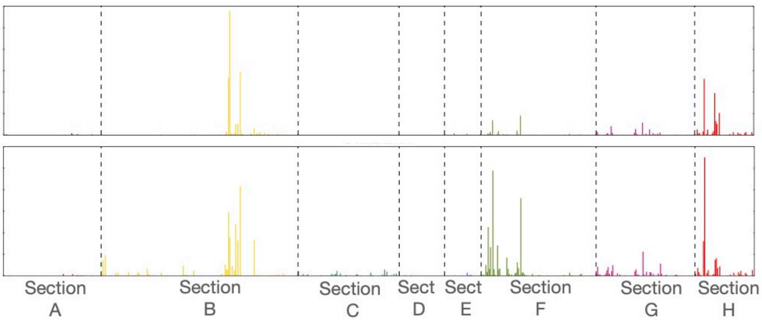

FIGURE 1. *TechSpectrum* for the IPC code *B60L* (top) and Toyota Motors (bottom) at IPC–subClass level

A *TechSpectrum* is a histogram, each bin of which corresponds to an specific IPC code, and it could be considered as an axe of a multidimensional space, namely the IPC space. The bin values would be in consequence coordinate values. Under this perspective, the visualization of technology, that is technical fields or tech companies represented by selected

[1]Bins are also known in other disciplines as intervals, classes or buckets

[2]Although the IPC has five levels – *Section*, *Class*, *SubClass*, *Group* and *SubGroup* – in the present work we have limited the analysis to the first four levels because the granularity at *SubGroup* level is too high and sensitive to the interpretation of the code definition.

sets of patents and its prior art citations, would be merely the representation of dots in the positions corresponding to the respective set of coordinates values, namely the bin values.

We have developed an analysis tool which generates two graphs, the *Technology Map* – ***TechMap*** – and the *Focused Technology Map* – ***F-TechMap*** – .

The first graph, the *TechMap* visualizes a given set of items, thus technical fields and tech companies, in a 2D plot in such a way that those related items they will appear closer between them that the unrelated ones.

The fact that our system has a high dimensionality, the dimension is 8, 131, 647 and 7545 at IPC–*Section*, *Class*, *subClass* and *Group* respectively[3], makes difficult a straightforward representation. An approach to visualize such high dimensional spaces is, first to reduce the dimensionality to two dimensions, and then generating a visualization in a 2D plane. For this task we have used *MultiDimensional Scaling – MDS –*, the downside of such an appreciable reduction of dimensionality is the introduction of a geometric distortion (Colange et al., 2019).

Note that for our aim of exploring proximity between technologies, *MDS* has the interesting feature of reducing the dimensionality to two dimensions preserving as much as possible the relative distances between the items to be visualized. In other words, items that are close in the *IPC* space stay close to each other, and those far stay far. On the other hand, the absolute positioning of the items within this visualization is meaningless, namely the fact that a given item is located in a particular position in the 2D plane provides no information, only the closeness or farness to any other item is meaningful (Borg and Groenen, 2005).

MDS visualizes the items in a 2D plane based on the distance information using a nonlinear dimensionality reduction, so the distances between the elements are the key information to gather. This information is technically provided as a distance o dissimilarity matrix.

The distance between items to visualize is computed with the *soft cosine*(Sidorov et al., 2014) algorithm, wherein the features are the IPC codes, and the similarity measure between features is computed applying a hierarchical similarity measure for hierarchical classification schemes(Caspersen et al., 2017).

Our second graph, the *Focused TechMap – F-TechMap –* is inspired on the work of Urpa and Anders (2019), and takes the shape of a circular–radial graph with the focus at the center of the circle and a set of radii, each corresponding to a single IPC code at a certain level of resolution. Once an item between the technical fields or companies present in the

[3]These are the number of IPC codes at each level on April 2022.

TechMap is chosen as to be the focus then every other item to be visualized is located at the computed distance on a *radius* corresponding to a given IPC-code.

The *F-TechMap* also uses the distances computed with the *soft cosine* algorithm, and it is a complementary graph to the *TechMaps* because it straight away uses computed distances avoiding the distance distortions produced by MDS in its projection on the 2D plane. The dynamic version of these two graphs will provide rich information about the evolution of companies technical developments and evolution of technical fields.

Two study cases are presented with some technical fields and tech companies. The data collection of the study cases is done using *PatStat–online* [4] an EPO's database in the field of patent intelligence and statistics [5]. A set of queries is executed in Patstat to collect the patents and citations of the technological fields or companies under study. In all our study cases patents are limited to the 1980–2015 range of years in order to have a large interval of time [6]. The patent documents collected are also constrained to be published by the USPTO[7] in order to avoid duplication.

The research question of the present work is what can we learn about tech companies development activities, and about technical fields evolution by visualizing patent classification information based maps.

The potential contribution of our work in the field of innovation management lies in the creation of a new graphical tool for the relative positioning of technologies (technical fields or tech companies) based on patent classification information, improving the perception and analysis of research and industrial activities of technology players. In the field of history of technology, our procedure contributes by facilitating the study of the evolution of specific technical fields and tech companies.

The rest of the paper is organized as follows: Section 2 discloses related works. Section 3 discloses the procedure to generate our graphical tool. Section 4 presents some study cases. Section 5 presents a discussion and the applications of our tool. Section 6 outlines future research and conclusions.

2. Related works

The visualization of technology has been used for long as a tool for analyzing technology and innovation (Geisler, 1962). The range of visualizations goes from relatively simple

[4]at https://www.epo.org/searching-for-patents/business/patstat.html

[5]PatStat contains bibliographical data relating to more than 100 million patent documents from leading industrialised and developing countries. It also includes the legal event data from more than 40 patent authorities contained in the EPO worldwide legal event data.

[6]Our procedure and PatStat allows different time intervals

[7]United States Patent and Trademark Office

graphs to visually rich and complex rendering such as Yoon and Magee for exploring technology opportunities (2018) and of Boyack et al. (2002) to visualize landscapes.

Patents are a fundamental component of the technology ecosystem, and patent information is employed extensively for technology visualization. For example, Yoon and Park(2004) have disclosed an analytical tool for high-technology trends which visualize patent networks based on text mining, or Liu and Zhu(2009) presented a system to visualize patent citations using web mining. Von Wartburg et al.(2005) have studied multistage patent citations to assess inventive progress. Moreover, maps of science or technology have been also produced by some authors such as Yan and Luo(2017) for measuring patent distances, Leydesdorff and Rafols(2013) using *Web-of-Science* categories, or Boyack et al.(2000) using patent citation networks for positioning specific patents on the landscapes.

Evaluating technological proximity based on patent information has been investigated by numerous authors. Bowen and Luo (2016) presented an overview of distance measurement for patent mapping. Simon and Sick (2016) disclosed the computation of technology distance is based on patent classification codes. And, Schoen et al. (2012) studied the evaluation of technological proximity based on patent information, specifically using co-classification.

Considering the above-mentioned *TechSpectrum* merely as a histogram makes possible to compute technological distances by computing histogram distances. Such methods are disclosed between other by Ma et al. (2010), Serratosa (2006), Strelkov (2008) or Werman (1985). However, in our case the histograms are very heterogeneous, especially some IPC *classes* have unimodal histograms in comparison with big tech corporations which have highly multimodal distributions, additionally the absolute figures of each frequency bin are also extremely different for some technologies represented by specific IPC codes and tech companies. These disparities in the histograms present problems resulting from the histogram normalization necessary for computing the distance based on histogram shape.

Considering the *TSpectrum* as a vector of the IPC space, the first solution is to use the *Euclidean* distance of the two vectors, that is the two *TSpectrum*s. However *Euclidean* distance is not scale–invariant and seems unsuitable for high-dimensional vectors(Simon and Sick, 2016).

An alternative is to compute the *Cosine* distance(Sidorov et al., 2014) because the result does not depend on the magnitudes of the vectors, but only on their angle. It remains the problem that not every *IPC* code is equidistant – orthogonal – because a given bin (an IPC code at a certain level) is usually much closer to other codes that belong to the same IPC *section*, *class* or *sub-class* than to those that belong to another[8]. An improvement to

[8]For example A61C is closer to A61B, A61D or A61L that to B61B, or A61B3 is closer to A61B5 that to A61C5.

that is the *Soft cosine* distance(Sidorov et al., 2014), which includes a factor among the different coordinates to take into account a certain *proximity* or *similarity* among them, meaning in our case a similarity factor between IPC codes.

Similarity in hierarchical structures such as a classification tree, conceptual taxonomy or ontology can be computed by using only the structural information of the tree, such as the concept similarity computation presented by Wu and Palmer (1994), which uses the number of nodes related to the concerned concepts. Li et al. (2003) discloses the computation of semantic similarity based on structural semantic information from a lexical taxonomy as a function of path length and depth between words. Caspersen et al. (2017) have disclosed a measure of similarity between labels in a hierarchical classification scheme for automatic classification.

The visualization of multidimensional data in a 2D plane was studied by numerous authors (Torgerson, 1952; Young, 1987; Borg and Groenen, 2005; Tenenbaum, Silva, and Langford, 2000; Lespinats et al., 2007). The basis of *MultiDimensional Scaling – MDS* – were disclosed by Torgerson in his seminal work(1952). Young(1987), and Borg and Groenen (2005) exposed some applications. An interesting use of *MDS* for exploring high-dimensional data is presented by Urpa and Anders(2019).

3. Technology Maps generation procedure

In our work a technology is defined as a set of patent documents assigned to a technical field or belonging to a tech company. The set of patents owned to an inventor, or a group of inventors could also be considered but this work is limited to tech fields and companies.

The patents are collected from *PatStat*, an EPO's database in the field of patent intelligence and statistics. A set of *SQL*–queries is executed in Patstat to collect the selected patents – and its prior art citations – of the technological fields or companies under study. The patent documents collected are constrained to first, the range of years from 1980 to 2015 in order to have a large interval of time, and second, to be published by the USPTO to avoid duplication[9].

The *TechMap* generation procedure is outlined in the following five steps:

Step 1: Data gathering.– First, we collect the set of patents published for the specific technical fields or tech companies to be visualized, then its cited prior art is also collected and mixed forming a final collection. In this final collection, the prior art is weighted to a certain percentage which experimentally we have set to 10%, the idea of this reduction is to keep the basic composition of the initial set of patents but enriching it with the prior art citations.

[9]PatStat allows bigger time intervals that 1985–2015 and virtually any patent office.

Step 2: *TechSpectrum* generation.– For each final collection the classification IPC codes are identified, and ordered in bins according to the IPC at the first four levels of classification resolution [10].

Step 3: Distance computation.– A matrix of distances between every and each item (technical fields or tech companies) to be visualized in the Tech Maps is computed using the *soft cosine* distance(Sidorov et al., 2014), which is the following formula:

$$SoftCosDist_{AB} = \frac{\sum_i^N \sum_j^N (S_{i_j} A_i * B_j)}{(\sqrt(\sum_i^N \sum_j^N S_{i_j}(A_i * A_j)) * (\sqrt(\sum_i^N \sum_j^N S_{i_j}(B_i * B_j))} \tag{1}$$

A and B represents vectors formed by bins of the final collections of two specific technologies (tech fields or companies). A_i and B_j – are the bin i and j of each vector, namely the IPC codes ordered according to the IPC at position i, and j of the two specific technologies.

For computing the *similarity* factor among each possible couple of IPC codes – S_{ij} – we have adapted the method proposed by Wu and Palmer(1994) to structural information of the hierarchical IPC scheme, namely the number of IPC levels between the path of each couple of IPC-codes C_i and C_j resulting the following formula:

$$Similarity(C_i, C_j) = \frac{N_{I_J}}{N_i + N_j + (2 * N_{I_J})} \tag{2}$$

N_i, N_j and N_{IJ} are the number of nodes on the path from C_i to C_{IJ}, from C_j to C_{ij}, and from C_{ij} to the *Root* node, where C_{IJ} is the least common IPC code of C_i and C_j.

In this manner, four distance matrices are formed with the *soft cosine* distances between every possible couple of items in the first four IPC levels. Then, we mix these matrices to form a *global* matrix [11]

$$GlobalDistance_{i_j} = K_S * d_{S_{ij}} + K_C * d_{C_{ij}} + K_{SC} * d_{SC_{ij}} + K_G * d_{G_{ij}} \tag{3}$$

K_S, K_C, K_{SC} and K_G, and d_C, d_S, d_{SC} and d_G are the mix factors and the *soft cosine* distances at IPC *Section*, *Class*, *SubClass* and *Group* respectively. We have experimentally set the parameters to K_S=1 , K_C=2, K_{SC}=3, and K_G=5.

[10] *Section*, *Class*, *SubClass* and *Group*.

[11] The rationale behind this operation is to catch the importance of a group of IPC codes at a level when at a sub-level the group of codes is populated but very distributed – flattened –. If we only compute and display at a specific IPC level the importance of the group of IPC codes could be minimized by this flattened effect

Step 4: *MDS* computation.– The dimensionality reduction is made at the *global* level by *MDS* using the *SKlearn* python library[12]. The *MDS* results in a collection of 2D coordinates corresponding to each item to be visualized.

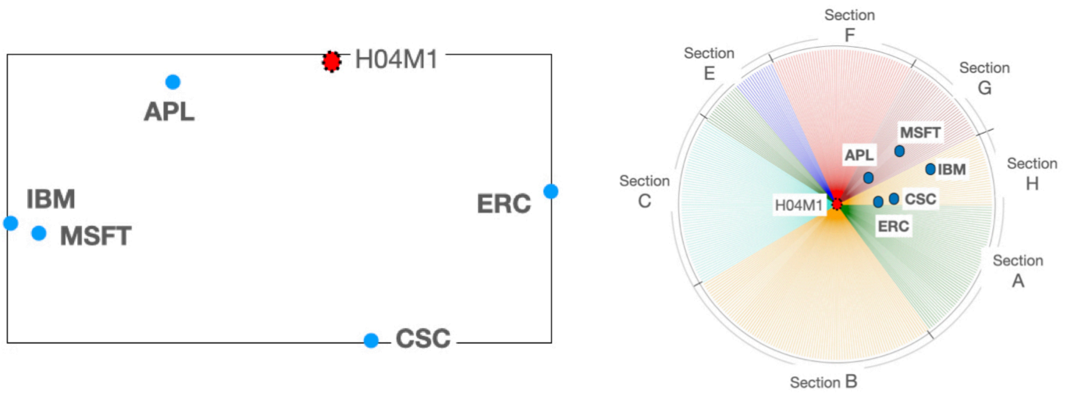

FIGURE 2. *Global TechMap* (left graph) with a technical field: telephone set – *H04M1* – and some computer and telecom companies: *Apple*, *Cisco*, *Ericsson*, *IBM* and *Microsoft*, and the corresponding *F-TechMap* (right graph) with focus at *H04M1*

Step 5: *TechMap* visualization.– The *MDS* results are visualized using the *MatPlotLib* python library(Devert, 2014). In the graphs, the tech companies are visualized as blue dots and the technical fields as red dots. An example of a *TechMap* with an IPC code[13] and some tech companies is shown in Figure: 2, left graph.

The *Focused TechMap* generation procedure is outlined in the following three steps:

Step 1: *TechMap* generation.– The global *TechMap* is generated and visualized for a given set of items (technical fields or tech companies) executing the *TechMap* generation procedure explained above.

Step 2: *Focus* selection.– Once the *TechMap* is generated, we select one of the visualized items to be the *focus* of the graph.

Step 3: *Focused TechMap* visualization.– The focus is drawn at the center of a circle wherein a set of radii is in turn plotted, this set contains a radius corresponding to each IPC code . Then, each item is drawn at its real computed distance to the focused item, on the radius corresponding to its main IPC code. The main IPC code for technical fields and

[12]At https://scikit-learn.org/stable/

[13]Appendix A contains the titles of all the IPC codes referenced in this paper

tech companies is the IPC code with the highest figures. An example of a *focused TechMap* with an IPC code and some tech companies is shown in Figure: 2, right graph.

4. Study cases

As above-mentioned, the data collection of the study cases is done using *SQL* queries in *PatStat* to gather the selected patents and its prior art citations, and the corresponding assigned IPC codes. The dimensionality reduction, and visualization is implemented with MDS, and the distance or dissimilarity matrix in all the study cases is computed using the *SoftCosine* distance.

For the study cases we have selected two technical fields, medical science – *A61* – and vehicle technology – *B60* –, and the following companies[14]: Apple (APL), Boston Scientific (BSc), Cisco (CSC), Ericsson (ERS), Ford Motor(FRD), G.E.-medical (GEm), General Motors (GMt), Hyundai Motors (Hyu), IBM (IBM), Medtronic (MDT), Nissan (Nss), Microsoft (MSFT), Nihon Kohden (NKH), Olympus-medical (OLYM), Omron-medical (OMR), Philips (PHLP), Shimadzu (SHI), Siemens HealthCare (SIEM), Toyota motors (TYT), Volkswagen (VW).

We have chosen the medical field for personal and academic interest since the origin of this work is in a research project in the biomedical engineering field. The second study case, vehicle technology was chosen for its general interest as a technology in important transformation at present and because it encompasses multidisciplinary properties. On the other hand, the set of tech firms was selected to have a heterogeneous group with big and mid-size corporations from different sectors and geographic origins.

4.1. **Medical Technologies.** In this study case we will use our tool to study some tech companies related with medical technologies.

Let's start exploring the location of our complete set of tech companies and the mayor technical fields where they are supposed to develop its activities. For that we will generate a *TechMap* with all the companies and medical science – *A61* –, vehicle technology – *B60* –, computation – *G06* – and telecommunications – *H04* –. Figure 3 shows this *Global TechMap*, wherein the red dots represent the IPC code and blue dots represent the tech companies.

Loking at Figure 3 appears three groups of companies: A first group around medical science, IPC code *A61* and formed by the companies *Boston Scientific, GE –medical–, Medtronic, Nihon Kohden, Olympus –medical–, Omron –medical–* and *Siemens HealthCare*

[14]In brackets are the shorhands used in the graphs for each company.

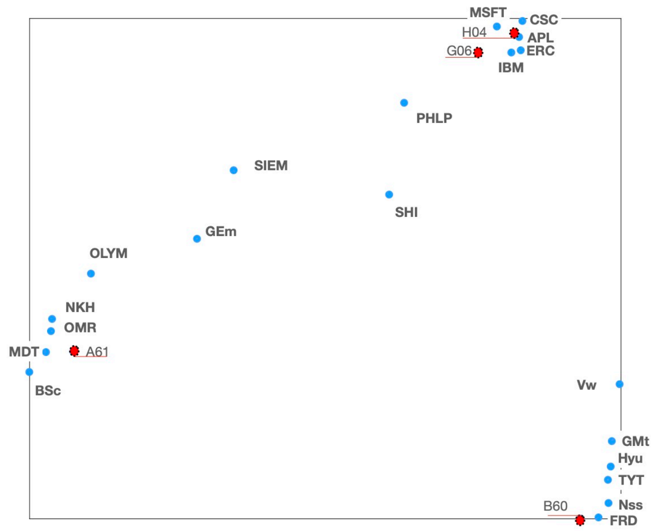

FIGURE 3. *Global TechMap* for some technical fields corresponding to IPC *classes* (red dots), and some tech companies (blue dots)

(see mid-left area in Figure 3). A second group close to computer science and telecommunication, IPC codes *G06* and *H04*, and formed by *Apple*, *Cisco*, *Ericsson*, *IBM* and *Microsoft* (see top-right area in Figure 3). A third group close to vehicle technologies, IPC code *B60* formed by the automotive companies *Ford Motors*, *GM*, *Hyundai*, *Nissan*, *Toyota* and *Volkswagen* (see bottom-right area in Figure 3). Finally, *Philips* and *Shimadzu* are located somewhere in between the "medical" and the "computer-telecom" groups.

Now, we will have a deeper look into the medical area going down a level from *A61* by generating a *TechMap* containing all the *A61* IPC *subclasses* and the companies around the medical area, *Boston Scientific, GE –medical–, Medtronic, Nihon Kohden, Olympus –medical–, Omron –medical–* and *Siemens HealthCare*(See Figure 4).

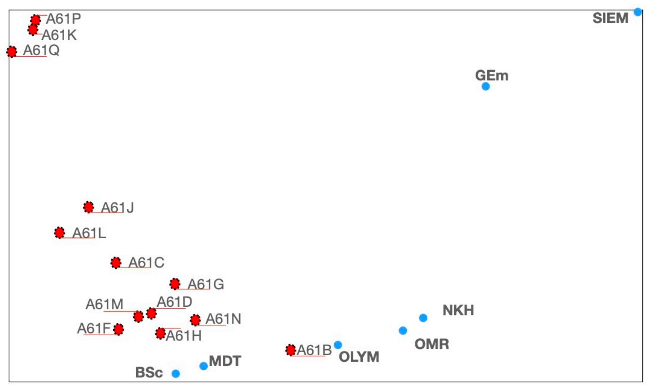

FIGURE 4. *Global TechMap* for medical companies and – *A61* – IPC *subclasses.*

It is interesting to highlight that by going deeper in the IPC, replacing a IPC class by its subclasses, or in a more general way replacing an IPC code at a level by its set of IPC sublevels, we are doing a sort of conceptual zoom rather than a zoom-in graphical operation, so that each item in the map is recomputed and located accordingly.

The new generated *TechMap* shows that the companies are grouped in a first set with *Boston Scientific* and *Medtronic*, *Olympus medical*, and *Nihon Kohden* and *Omron medical*, and a second set with *Siemens Healthcare* and *GE medical* (See Figure 4 at the bottom and the top-right respectively). The first group is located around diagnosis techniques – *A61B* –. It appears from Figure 4 that *Nihon Kohden*, *Omron* and *Olympus* are particularly close to this technique, whereas *Boston Scientific* and *Medtronic* are also located close to *A61N*, *A61H* and *A61F*.

Let's zoom-in around *Nihon Kohden*, *Omron* and *Olympus* and *A61B*. This is done going one IPC level down from *A61B*, so generating a new *TechMap* with the IPC *A61B groups*[15] and the three tech companies (See Figure 5).

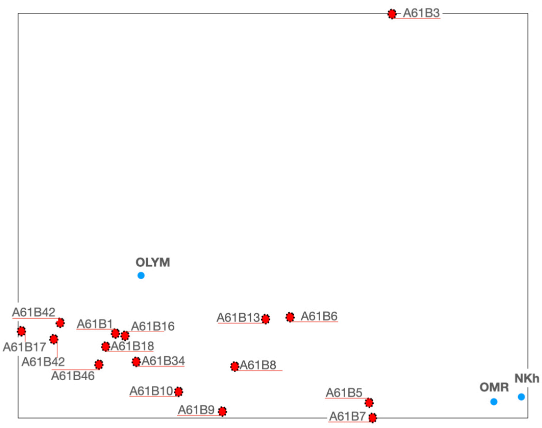

FIGURE 5. *Global TechMap* with all *A61B* groups and *Nihon Kohden*, *Omron medical* and *Olympus medical*

Two observations can be made from this *TechMap*: firstly, *Nihon Kohden* and *Omron* are grouped and close to *A61B5* and *A61B7*, whereas *Olympus* is close to a group of IPC *groups* formed by *A61B1*, *A61B16* and *A61B18*. Secondly, some IPC *groups* appear in clusters, one with *A61B5* and *A61B7*, and another cluster with *A61B1*, *A61B18* and *A61B16*.

[15] *A61B* has the following IPC *goups*: *A61B1*, *A61B3*, *A61B3*, *A61B6*, *A61B7*, *A61B8*, *A61B9*, *A61B10*, *A61B13*, *A61B16*, *A61B17*, *A61B18*, *A61B18*, *A61B34*, *A61B42*, *A61B46*, *A61B50*, *A61B90*

Olympus medical appears in Figure 5 close to a certain number of technical fields, so we will now use the *focused TechMap* tool to perceive without any distance distortion the closest and therefore the most important technical fields for that company. The *focused TechMaps* with *focus* on *Olympus medical* shows that endoscopes – *A61B1* – are by far the main technical field of its developments followed by computer-aided surgery – *A61B34*– (see in Figure 6).

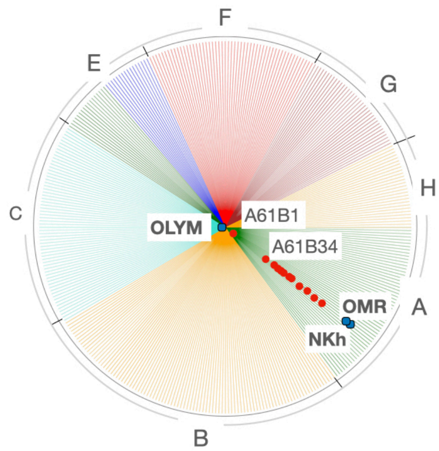

FIGURE 6. *Focused TechMaps* with focus on *Olympus medical*

4.2. **Automotive Technologies: Electric vehicles.** In this study case, we will study the evolution of the companies which are located in the *TechMap* of Figure 3 close to vehicle technology, namely *Ford*, *General Motors*, *Hyundai Motors*, *Nissan*, *Toyota Motors* and *Volkswagen* (see at the bottom right of Figure 3). We will study these companies in relation to electric vehicles. Moreover, in this case the data will be visualized in time-lapses of 5 years (from 1985 to 2015) to have a dynamic perception.

Electric vehicle technology is covered by the IPC subclass *B60L*. So, we will generate a first *TechMap* containing this subclass and our set of motor companies for the whole time interval (from 1985 to 2015) to have an overview of the relative positioning of the companies to the Electric vehicles technology.

The *TechMap* shows *Volkswagen* out of the group formed by the rest of the companies, and additionally it seems that the more active companies are *Hyundai*, *Toyota* and *Nissan* (see Figure 7).

Now, we will visualize its evolution by computing a series of *TechMaps* for five years time interval. Figure 8 shows (from right to left, and from top to bottom) the six *TechMaps* for 1985 to 1989, 1990 to 1994, 1995 to 1999, 2000 to 2004, 2005 to 2009, and 2010 to 2015.

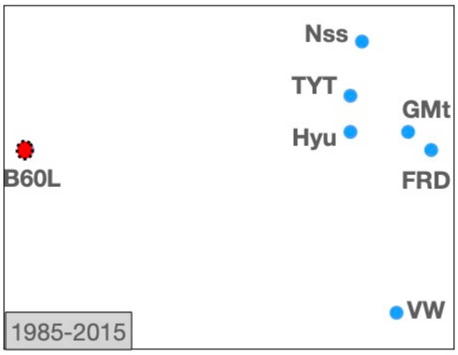

FIGURE 7. *TechMaps* **from 1985 to 2015** for *B60L* IPC sub-class and the motor companies: *Ford*, *General Motors*, *Hyundai Motors*, *Nissan*, *Toyota Motors* and *Volkswagen* .

FIGURE 8. *TechMaps* evolution for 5-year intervals from 1985 to 2015 containing the *B60L* IPC sub-class and the motor companies: *Ford*, *General Motors*, *Hyundai Motors*, *Nissan*, *Toyota Motors* and *Volkswagen* .

It can be observed that, it is in the late 90s when some car companies started to approximate the *B60L* subclass. It appears that *Hyundai Motors* is far from this field until the beginning of 2000 when it joined the group of companies more active in electric vehicle technology. It also seems clear that *Nissan* stays from the late 90s as one of the companies more active in the field. It is also interesting to note how *Volkswagen* remains as one of the companies less active in this field until the 2010s.

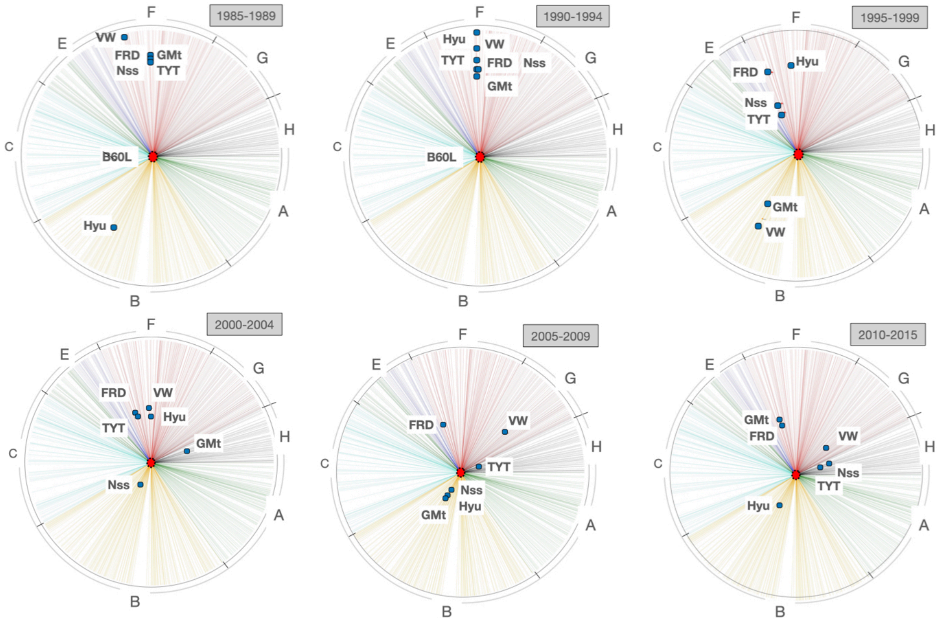

FIGURE 9. *Focused TechMaps* with focus in *B60L*. Evolution from 1985 to 2015 at 5-years intervals

We will now use the *F-TechMap* to *focus* on *B60L* for improving our perception of tech companies around electric vehicle technologies. As above-mentioned [16] the visualized items, in this case the motor car companies, will be located at actual distances from the *focus*, the electric vehicle technology *B60L*, and each item will be placed on the radius corresponding to most assigned IPC codes to it.

The *Focused TechMaps* confirm that from the interval 1995–1999, the car companies started to be more and more active in the field, pointing out *Toyota* and *Nissan*, and from the interval 2000–2005, also *Hyundai* (see in Figure 9 top right and bottom left graphs).

An interesting fact shown in these graphs is the change of the most important IPC section for some companies. In Figure 9 the three top graphs (1985 to 1999) place the companies on radii corresponding to sections B and F, which correspond to transporting and mechanical engineering technologies respectively. These technical fields are traditionally considered as car technologies, whereas from 2000 (see in Figure 9 the three bottom graphs), some of these companies started to be located on electricity-physics technologies.

[16] See step 3 of *Focused TechMap* methodology on page 8

Note that by the interval 2010-2015 three companies are located in sections H or section G. In fact, *Nissan* and *Toyota* are both located in section H (electricity), specifically in the subclass assigned to battery technology – *H01M* –. On the other hand, *Volkswagen* is located in section G (physics), specifically in the subclass assigned to digital data processing – *G06F* –. Somehow these car companies are emerging as electrical technology developers in addition to, or above mechanical technology.

5. Discussion

The visualization of proximity between technical fields and tech companies with respect to each other furnishes an insight in industrial and research developments. Such visualization provides information about firms know-how, which facilitates potential collaborations and, technology opportunities and weakness. From this perspective our work presents two complementary graphical tools for studying the relationship of technical fields and tech companies, and its evolution.

The information provided by our tool can help technology decision makers to understand competitors. For example, the evolution of motor car companies towards electricity technology can be explored and better understood by generating the *TechMaps* and then focusing on specific IPC codes or companies. Such graphs could also help to perceive the kind of knowledge that a specific technology gathers at a certain moment. Furthermore, the identification of these trends can be of great interest to researchers in the area of the history of technology.

The relative positioning of tech companies can be used to discover know-how or to explore complementarities between companies for potential merge and acquisition, or research collaborations. See for example the position of *Apple* in the *TechMap* and the *focused TechMap* of Figure 2 as a telephone set developer together with *Ericsson* in relation to *IBM* and *Microsoft*, two computing companies in principle similar to *Apple*.

It is important to highlight that our approach is to consider technical fields and companies (both defined by sets of IPC codes) as similar bodies, and they are represented by the same data structure, the *TechSpectrum*, namely the histogram of the IPC codes assigned to a set of patents and its citations mixed in a specific proportion, and ordered according to the IPC. This is a major difference with respect to the previous studies on visualization of technical fields and companies.

This conception of technical fields and tech companies as similar bodies of technology bring important implications in the processing and visualization of technology. First, the processing is the same for technical fields and companies, so changes in time in both are tacking equally into account providing an overview on technical change as a whole. Second, observations for technical fields can be extend to companies.

Our *TechMap* has some distinctive features from traditional technology maps or landscapes. In *TechMap* the items, technical fields or companies, are located relative to each other, there is not absolute positioning. In previous maps or landscapes, the location of technology items is computed and coordinate data are assigned to it, so that if you add or remove an item, it is added or removed to the map but the rest of it remains unchanged, whereas in *TechMap*s removing or adding a technology item results in a computation of the new position of each item. Another distinctive feature is the capability to visualize tech companies in relation to technical fields at different levels of conceptual resolution by going deeper in the IPC classification levels. This operation is a sort of conceptual zoom rather than a zoom-in graphical operation as is usually the case in-known technology maps or landscapes such us the multiresolution zooming of Boyack et al (2000).

The presence of clusters of technical fields within our *TechMap*s can show up the existence of some commonalities which can be useful for researchers and innovation managers to better understand these technical fields. See for example in Figure 4 how the IPC *subclass A61P*, *A61k* and *A61Q* appear grouped. Or the, not obvious proximity between measuring systems for diagnosis – *A61B5* – and instruments for auscultation – *A61B7* – shown in Figure 5. It is interesting to highlight the exploration possibilities open by visualizing with our two graphical tools tech companies in relation to technical fields, such as in Figure 2 where *IBM* and *Microsoft* appear grouped when they are positioned within a set of companies and the technical field of telephone sets – *H04M1* –.

6. Summary and Future Work

In this work we introduced a new graphical tool based on patent classification information to study technical fields represented by an IPC code (or a combination of codes) and tech companies. Each item, technical field or company is characterized by a weighted mix of the classified (or assigned) patents and its cited prior art.

The first graph, the technology map – *TechMap* –, visualizes a set of given technologies, technical fields and tech companies positioning them in relation to each other. Our second graph, the focused TechMap generates a visualization of the given technologies in relation to a selected one, the focus, which is located at the centre of a circle with drawn radii corresponding to every IPC code. In this graph, the items are positioned at the computed distances from the focus, and on the radius representing its IPC code with the highest figures.

Although we have illustrated our tool with two study cases, namely medical and automotive technology, our tool has also been tested in a variety of fields such as heartbeat monitoring, 3D printing, hybrid vehicles, electronic devices, computer graphics or telecommunications to test for the ecological validity of our tool. So far, the authors found that this is a representative view in terms of ecological validity except for the chemistry field which needs further dedicated testing due to its construct nature of how patents are triaged and presented. Concerning tech companies, we have tested the tool with numerous firms and we have found that their use in our tool is constrained to firms active in patenting.

In the field of history of technology, our graphical tool contributes by facilitating the study of the evolution of specific technical fields, and to trace the divergence or convergence of tech companies. The contribution of the present paper lies in the creation of two new and complementary graphs based on patent classification information. The relative positioning of technologies (technical fields or tech companies) helps to better identify those that have some techniques in common because they appear close in the graphs, as well as to improve the understanding of the technical developments or trends in firms as was illustrated for motor car companies moving from pure mechanics to electrical technologies.

At present we are developing some algorithms for automatic clustering of items within *TechMaps*. Moreover, we foresee doing predictive modelling of the dynamics of the technologies to anticipate its evolution, especially for tech companies.

Further research will be oriented to compute *TechMaps* with smaller time intervals in order to have more time resolution in the evolution perception to present animated versions of the visualizations and to investigate the modelling of trends.

We will explore the improvement of the similarity matrix using text similarity and citation network analysis of the patents classified in the respective IPC codes.

References

Alstott, J. et al. (2017). "Mapping technology space by normalizing patent networks". In: *Scientometrics* 110.1. doi:`https://doi.org/10.1007/s11192-016-2107-y`, pp. 443–479.

Borg, I. and P.J. Groenen (2005). *Modern multidimensional scaling: Theory and applications.* Springer.

Bowen, Y. and J. Luo (2016). "Measuring technological distance for patent mapping." In: *Journal of the Association for Information Science and Technology* 68, pp. 423–437.

Boyack, K.W., B.N. Wylie, and G.S. Davidson (2002). "Domain Visualization Using VxInsight® for Science and Technology Management". In: *Journal of the American Society for Information Science and Technology (JASIST)* 53.9. doi:`https://doi.org/10.1002/asi.10066`, pp. 764–774.

Boyack, K.W., B.N. Wylie, G.S. Davidson, and D.K. Johnson (2000). "Analysis of Patent Databases Using VxInsight". In: *9th International Conference on Information and Knowledge Management (CIKM 2000).*

Caspersen, K.M. et al. (2017). "A Hierarchical Tree Distance Measure for Classification". In: *Proceedings of the 6th International Conference on Pattern Recognition Applications and Methods* 1. doi:`https://doi.org/10.5220/0006198505020509`, pp. 502–509.

Colange, B. et al. (2019). "Interpreting Distortions in Dimensionality Reduction by Superimposing Neighbourhood Graphs". In: *2019 IEEE Visualization Conference (VIS).* doi:`https://doi.org/10.1109/VISUAL.2019.8933568`, pp. 211–215.

Devert, A. (2014). *Matplotlib Plotting Cookbook.* PACKT.

Geisler, E. (1962). *The Metrics of Science and Technology.* Quorum Books, Greenwood Publishing Group Inc.

Jaffe, A.B. (1986). "Technological opportunity and spillovers of R&D: evidence from firms' patents, profits and market value." In: *American Economic Review.*

Lespinats, S. et al. (2007). "DD-HDS: A method for visualiza-tion and exploration of high-dimensional data". In: *IEEE Transactions on Neural Networks* 18.5, pp. 1265–1279.

Leydesdorff, L., S. Carley, and I. Rafols (2013). "Global maps of science based on the new Web-of-Science categories". In: *Scientometrics* 94. doi:`https://doi.org/10.1007/s11192-012-0784-8`, pp. 589–593.

Li, Y., Z.A. Bandar, and D. McLean (2003). "An Approach for Measuring Semantic Similarity between Words Using Multiple Information Sources". In: *IEEE Transactions on Knowledge and Data Engineering* 15.4. doi:`https://doi.org/10.1109/TKDE.2003.1209005`, pp. 871–882.

Liu, Z. and D. Zhu (2009). "Web Mining based Patent Analysis and Citation Visualization". In: *Proceedings.* doi:`10.1109/WMWA.2009.33`, pp. 19–23.

Ma, Y., X. Gu, and Y. Wang (2010). "Histogram similarity measure using variable bin size distance". In: *Computer Vision and Image Understanding* 114, pp. 981–989.

Okubo, Y. (2017). "Oikawa, K. (2017). Inter-firm technological proximity and knowledge spillovers." In: *Tokyo Center for Economic Research (TCER)* Paper-E114.

Perez-Molina, E. (2018). "The role of patent citations as a footprint of technology". In: *Journal of the Association for Information Science and Technology* 69.3. doi:`https://doi.org/10.1002/asi.23979`, pp. 610–618.

Schoen, A. et al. (2012). "The Network Structure of Technological Developments; Technological Distance as a Walk on the Technology Map". In: *Science and Technology Indicators Conference.*

Serratosa, F. and A. Sanfeliu (2006). "Signatures versus histograms: Definitions, distances and algorithms". In: *Pattern Recognition* 39, pp. 921–934.

Sidorov, G. et al. (2014). "Soft Similarity and Soft Cosine Measure: Similarity of Features in Vector Space Model". In: *Computación y Sistemas* 18, pp. 491–504.

Simon, H. and N. Sick (2016). "Technological distance measures: new perspectives on nearby and far away". In: *Scientometrics* 107, pp. 1299–1320.

Strelkov, V.V. (2008). "A new similarity measure for histogram comparison and its application in time series analysis". In: *Pattern Recognition Letters* 29, pp. 1768–1774.

Tenenbaum, J.B., V. de Silva, and J.C. Langford (2000). “A Global Geometric Framework-for Nonlinear DimensionalityReduction”. In: *Science* 290, pp. 2319–2323.

Torgerson, W.R. (1952). “Multidimensional Scaling: Theory and Method”. In: *Psichometrika* 17.4, pp. 401–419.

Urpa, L.M. and S. Anders (2019). “Focused multidimensional scaling: interactive visualization for exploration of high-dimensional data”. In: *BMC bioinformatics* 20.1, pp. 1–8.

vonWartburg, I., T. Teichert, and K. Rost (2005). “Inventive progress measured by multi-stage patent citation analysis”. In: *Research Policy* 34, pp. 1591–1607.

Werman, M., s. Peleg, and A. Rosenfeld (1985). “A Distance Metric for Multidimensional Histograms”. In: *Computer Vision, Graphics, and Image Processing* 32, pp. 328–336.

Woerter, M. (2012). “Technology proximity between firms and universities and technology transfer”. In: *The Journal of Technology Transfer* 37. doi:`https://doi.org/10.1007/s10961-011-9207-x`, pp. 828–866.

Wu, Z. and M. Palmer (1994). “Verb Semantics and Lexical Selection”. In: *Proceedings of the 32nd annual meeting on Association for Computational Linguistics*. doi:`https://doi.org/10.3115/981732.981751`, pp. 133–138.

Yan, B. and J. Luo (2017). “Measuring Technological Distance for Patent Mapping”. In: *JASIST* 68.2. doi:`https://doi.org/10.1002/asi.23664`, pp. 423–437.

Yoon, B. and C.L. Magee (2018). “Exploring technology opportunities by visualizing patent information based on generative topographic mapping and link prediction”. In: *Technological Forecasting and Social Change* 132. doi:`https://doi.org/10.1016/j.techfore.2018.01.019`, pp. 105–117.

Yoon, B. and Y. Park (2004). “A text-mining-based patent network: Analytical tool for high-technology trend”. In: *Journal of High Technology Management Research* 15. doi:`https://doi.org/10.1016/j.hitech.2003.09.003`, pp. 37–50.

Young, F.W. (1987). *Multidimensional scaling. History, Theory and Applications*. LEA publishers.

Appendix A: Exact titles of cited IPC classification codes

A61 : MEDICAL OR VETERINARY SCIENCE; HYGIENE
A61B : DIAGNOSIS; SURGERY; IDENTIFICATION
A61C : DENTISTRY; APPARATUS OR METHODS FOR ORAL OR DENTAL HYGIENE
A61D : VETERINARY INSTRUMENTS, IMPLEMENTS, TOOLS, OR METHODS
A61F : FILTERS IMPLANTABLE INTO BLOOD VESSELS; PROSTHESES; DEVICES PROVIDING PATENCY TO, OR PREVENTING COLLAPSING OF, TUBULAR STRUCTURES OF THE BODY
A61G : TRANSPORT, PERSONAL CONVEYANCES, OR ACCOMMODATION SPECIALLY ADAPTED FOR PATIENTS OR DISABLED PERSONS
A61H : PHYSICAL THERAPY APPARATUS
A61J : CONTAINERS SPECIALLY ADAPTED FOR MEDICAL OR PHARMACEUTICAL PURPOSES; DEVICES OR METHODS SPECIALLY ADAPTED FOR BRINGING PHARMACEUTICAL PRODUCTS INTO PARTICULAR PHYSICAL OR ADMINISTERING FORMS; DEVICES FOR ADMINISTERING FOOD OR MEDICINES ORALLY; BABY COMFORTERS; DEVICES FOR RECEIVING SPITTLE
A61K : PREPARATIONS FOR MEDICAL, DENTAL, OR TOILET PURPOSES
A61L : METHODS OR APPARATUS FOR STERILISING MATERIALS OR OBJECTS IN GENERAL; DISINFECTION, STERILISATION, OR DEODORISATION OF AIR; CHEMICAL ASPECTS OF BANDAGES, DRESSINGS, ABSORBENT PADS, OR SURGICAL ARTICLES; MATERIALS FOR BANDAGES, DRESSINGS, ABSORBENT PADS, OR SURGICAL ARTICLES
A61M : DEVICES FOR INTRODUCING MEDIA INTO, OR ONTO, THE BODY
A61N : ELECTROTHERAPY; MAGNETOTHERAPY; RADIATION THERAPY; ULTRASOUND THERAPY
A61P : SPECIFIC THERAPEUTIC ACTIVITY OF CHEMICAL COMPOUNDS OR MEDICINAL PREPARATIONS
A61Q : SPECIFIC USE OF COSMETICS OR SIMILAR TOILET PREPARATIONS
B60: VEHICLES IN GENERAL
B60L: PROPULSION OF ELECTRICALLY-PROPELLED VEHICLES; SUPPLYING ELECTRIC POWER FOR AUXILIARY EQUIPMENT OF ELECTRICALLY-PROPELLED VEHICLES ; ELECTRODYNAMIC BRAKE SYSTEMS FOR VEHICLES IN GENERAL; MAGNETIC SUSPENSION OR LEVITATION FOR VEHICLES; MONITORING OPERATING VARIABLES OF ELECTRICALLY-PROPELLED VEHICLES; ELECTRIC SAFETY DEVICES FOR ELECTRICALLY-PROPELLED VEHICLES
G06: COMPUTING; CALCULATING OR COUNTING
H01: BASIC ELECTRIC ELEMENTS
H01M: PROCESSES OR MEANS, e.g. BATTERIES, FOR THE DIRECT CONVERSION OF CHEMICAL ENERGY INTO ELECTRICAL ENERGY
H04: ELECTRIC COMMUNICATION TECHNIQUE
H04M: TELEPHONIC COMMUNICATION
H04M1: Substation equipment, e.g. for use by subscribers