Patent based analysis of Technology Maps for Innovation Management and Research Planning

E. Sani[1,*], E. Ruffaldi[1], M.Bergamasco[1]

[1] Scuola Superiore Sant'Anna, Pisa, Italy

* Corresponding author: e.sani@sssup.it

Abstract: Innovation management and research directions planning are challenged by the understanding of the technology space in which an entity is placed. In particular, it is necessary to understand the relationships of the patent portfolio with the respect to competitors. Patent databases provide useful information for exploring a given technology area, but only when they are being properly analyzed are able to provide fruitful insights. The approach of technology trajectories can be extended toward the creation of an intuitive technology map. This paper presents an interactive approach to the exploration of technology space based on patents. Patents are being analyzed using the citation graphs generating a technology map that presents the technology topics on a landscape, displaying the patent density and the main actors in the area. The map is interactively displayed allowing the user to focus on specific patents or companies. Finally, a specific case study is provided presenting the application of this approach focusing on the area of rehabilitation.

Keywords: patent; citation, technology mapping; research planning; patinformatics

I- Introduction

Nowadays if a firm wants to lead the market and take advantage of its positioning, it has to exploit the same information about technology that are shared among other competitors. Private agents standing out the market and profit-seeking allocate specific resources for exploring and developing new products or technologies they suppose to be promising. If they succeed in doing it, their production costs, market position, competitors, and clients will be affected by such innovative techniques. In some cases, the way things go has negative sign and the consequences may be severe. As well known, the New Product Development process is a time and cost consuming component and there is a lot of sense in considering the information technology as an enabling tool for improving the effectiveness of management operations and management strategic decision. Indeed, from a business point of view, research is a risky and expensive activity, that may or may not provide exploitable results; it is undertaken especially when it is not strategically wise to achieve the same technology with different policies (licensing, cross-licensing, etc.). The main result of industrial research is the patent, being the official and formal way of innovation protection. The patent portfolio represents one of the key assets especially for high technology companies, and it can be adopted for several business goals (e.g. keeping market position, supporting profitable agreements, protecting current, future products). In the context of patents and the related licensing, exhaustive surveys are mainly focused on the adoption of innovative technologies, and the related economic impact, without identifying the primary causes of such innovation (Rosenberg, 1982) or identifying critical path analysis. It is indeed interesting to study possible paths that can be generated once having analysed patent samples and their relationships in a given area. Following this trend, recent studies try to verify the theory of invention (Fleming, Sorenson 2001) following the Kauffman's work (Kauffman, 1993) based on evolutionary biology. Kauffman focused his analysis considering the role of complexity in adaptive systems and assuming the existence of a landscape in which superior organisms can seek superior levels of biological fitness. This metaphor can be interpreted in the sense that, the stronger is an innovative technology on a market, the higher is the likelihood to establish fruitful collaborations and agreements with third parties. Along the lines of this, further studies demonstrate that technologies follow a sort of 'technological life-cycle' (Abernathy 1978, Anderson 1990) and further else that technologies go beyond different periods of equilibrium broken in unsteady ones (Eldredge, Gould 1972).

This paper focuses on technology innovation management and research directions planning as an interesting tool for a strategic understanding of the technology space of given entities in a given area. Following on from a brief literature review of patent analysis, research planning, roadmapping and technology trajectories, this work focuses on describing the methodology adopted. A case study is represented showing the application of the technology mapping process on specific fields of technologies related to rehabilitation.

II- Research background

II.i Patent analysis applications

The tools for Intellectual Property analysis, open to business R&D departments as well as to University Technology Transfer Offices (TTOs), improved along time both in qualitative and quantitative aspects. These tools allow to retrieve patents and search among the relationships of patents, but they do not provide a detailed content and cross referenced based search. The opportunity offered by more intelligent text analysis algorithm, integrated with domain knowledge bases, allows the creation of new tools for the analysis of IP material. These new tools, based on citation networks and semantic analysis, let discover relationships between patents and new patents previously hidden. Some analysis tools are already available for helping and structuring the clear placement of the company with the respect to the outside market as well as for identifying future directions, based on the information sourcing in public database of patents (Jenkins, 2001), although there is space for more interesting analysis. As pointed out by other researchers, since inventors probably limit their patent applications to their more successful inventions, patents presumably represent only the higher tops on conceivable and potential landscapes (Fleming, 2001). Nevertheless, the number of citations a patent receives is strictly and positively correlated with its technological relevance (Hall, 2000), (Albert 1991) and may also be tightly connected with the social value, representing a measure of the inventive usefulness across heterogeneous and different technologies fields (Fleming, 2004). The proposed analysis focuses on patents and patents citations as a measurement of the current trends of a specific technological scientific innovation useful also for analyzing specific technologies niches (Sani, 2007).

II.ii Technology trajectories as a complex informative landscape

The evolution of technologies in a given area is an important step in the understanding of a company placement and in the identification of possible future directions. In particular, the concept of technology trajectories (Dosi 1982) has been introduced to formalize and present the major lines of the evolution of technologies; a technology trajectory can indeed be represented as a path that, among the various possible paths of technological solutions, has been successful in a given area along time. This type of analysis has been successfully applied to a variety of fields, and it has been primarily based on the use of patent citations (Helo, 2003), (Verspagen 2007), (Martinelli 2008). The quantitative results provided by this type of analysis are shown in the form of graphs with bifurcation when there are several technological alternatives. In this way it is interesting to identify and study solutions that have been discarded or others that could be promising. The insights provided by technology trajectories are extremely useful in the understanding of the general trend of a technological field, but for a deeper analysis it is necessary to use more advanced types of visualizations that can provide more complete and detailed scenarios.

In particular an interesting technique for presenting a domain of data or a complex data set is the landscape visualization in which abstract data is represented as a map (Boyack 2002). Specifically, technology maps are able to represent the major relationships between technologies and the positions of different competitors in this landscape (Rinne 2004). These maps can be constructed based on textual information, as in data mining and topic maps (Boyack, 2001), or based on the citation associated to documents (Brandes, 2002). The adoption of landscape requires anyway some care because 3D visualization is not guaranteed to provide a better understanding of a problem (Tory, 2007). The other important aspect that should be taken into account in the landscape visualization is the time component; it represents the evolution of the patents and technology along time (Huang Z. et al. 2003).

This paper bases the innovation management approach on the creation of technology maps that allow to interactively understand the position of a company with the respect to current trend of technologies, as they can be obtained from patent information.

III- Methodology

The proposed approach uses the information extracted from a given patent set to create a technology map that can be explored for understanding the technology landscape along time. The phases from the patent data to the map are the following:

1) Construction of the patent set
2) Computation of the citation graph
3) 2D Layout of the citation graph
4) Analysis of areas and portfolios
5) Creation of the landscape

The patent set is currently created during a first phase of the analysis, although it could be performed incrementally while the map is being created. This creation phase starts from a elementary query in the patent database, or simply by selecting a group of well known patents. This initial set is extended first by extending the search to the International Patent Class (IPC) subclasses of this initial patent set, eventually using specific keywords contained in the abstract and in the body of the initial set. Finally, the set is extended by recursively traversing the citations network of the patents, using various criteria as a stopping condition. Once the patent set has been obtained, it is important to normalize the contained data and to aggregate the duplicate entries. In many cases, indeed, the same patent is found in multiple versions or databases, or the broad search provides results in patents belonging to the same family. After that, it is assumed to not modify the patent set, and for this reason, every patent is associated with an index in the patent set. In Information Retrieval terminology the patent is equivalent to a document and the patent set is the database.

The method applied for the analysis and the visualization of patent data is twofold. One part of the analysis involves the use of the citation graph, that is the network of citations among patents, while the second part is based on text mining techniques applied on the patent text. The citation graph can be considered as a flow of information from older patents to newest ones; such a flow is reversed with the respect to the citations direction. The graph is represented by an adjacency matrix C, in which the element i,j is set to one when the information flows from the patent at position i to the patent at position j. From the matrix C is possible to build the reachability matrix R, that informs us if a patent i is connected some way to a patent j. From this matrix it is possible to identify the disjoint components of the patent set, that are single patents or group of patents that are separate from others. Such components cannot be connected using the citation analysis. The integration of text based techniques allow to compute distances respect these components.

Finally, we distinguish patent depending on their information flow: a source is a patent that is not receiving information while a sink is a patent that only receives information. The citation graph analysis allows to capture the major structures in the citation network, and for this purpose we are adopting the algorithm of Search Path Link Count (SPLC) as introduced by (Hummond and Doreian, 1989) and used by (Verspagen 2007) for patent analysis. The SPLC of an edge in the citation graph is the number of shortest paths from all sources to all sinks that pass over a given edge: this is equivalent to identifying the most common citations among all the ones in the citation space. The SPLC is typically used for computing the main path in the trajectory that is the most relevant trend in the network, based on this type of weighting. The objective of this work, anyway, is the identification of the different technological trends. The approach followed in this work is based on the iterative computation of the main paths, at every step a main path is computed and then all the associated weights are removed from the network. The resulting paths are representative of different topics. Each topic can be analyzed by building a sub-network of the patents relative associated to the main path. This sub-network can be built by propagating the patents involved in the network following the flow of information up to a certain level.

The trajectory path provides the fundamental structure of the citation graph and it is used for creating the core layout of the patent set, using force directed algorithms like Fruchterman-Reingold. The 2D layout is being transformed into a 3D surface mapping the density of patents in a given area to the height. The computation of the height from the single patent data is performed using Radial Basis Functions (RBF): ever patent is a center of a template surface, typically a Gaussian, and the surface is obtained by the superimposition of all these surfaces. The alternative to this process is an interpolation based on triangulation of the point, that for unevenly spaced points like the patents in the layout generates surfaces that are not sufficiently smooth for the interactive exploration. The color of the landscape represent also the density supporting the information provided by the height. The metaphor of the landscape is augmented with the display of the trajectories as curved lines over the landscape, showing the different relevant regions.

The relationship between players in a given topic is obtained by analyzing the assignee field of the patent data. In the cleaning phase the names of the assignees are normalized for making sure of selecting research institution and companies. In this way each assignee identifies a different cluster of the citation graph, and in addition it is possible to use the 2D layout information for measure the distance of a patent portfolio respect the main path. Assuming that the top path is a good representation for the trend in an area it is possible to evaluate the distance of a firm respect the path, and estimate this trend along time (Martinelli 2008).

In terms of implementation the patent set is obtained from esp@cenet and USPTO and later processed with MATLAB. The layout is performed using Force Directed Placement initially based on Graphviz and then re-implemented for exploiting the parallelism of Graphical Processing Unit (GPU). The interactive visualization in 3D is based on MATLAB, allowing a fast prototyping of the interaction.

IV- Case study

This paper talks about technology mapping on specific fields of technologies related to rehabilitation. In particular we started from a preliminary query on patents in the IPC class A61 containing the word prefix *rehab* in the title and in the abstract, from US and EU databases. This starting set, after duplication removal contained 622 patents. The second phase of the data extraction has been based on the recursive retrieval of patent following the citation graphs, giving 10463 patents after two level recursion, and 3851 with one level recursion. This investigation is based on the set of 10463 patents that are organized in 31 disconnected components. The largest component contains the 97% of the patents, while the others contain between 2 and 36 patents. In this analysis we are going to focus on the biggest component of 10212 patents. In terms of time the period covered by the patent set is from 1967 to 2005.

The major trajectories computed with SPLC are:

1. from simple hearing aids to rehabilitation of hearing (1973-2000)

2. from basic knee brace to anatomically and neurophysilogically designed knee braces (1976-2003)

3. electric surgical instruments (1975-2004)

For example the first trajectory in the main component of this set is associated to the following patents:

Patent	Year	Title
US3764748	1973	IMPLANTED HEARING AIDS
US3870832	1975	IMPLANTABLE ELECTROMAGNETIC HEARING AID
US4352960	1982	Magnetic transcutaneous mount for external device of an associated implant
US4606329	1986	Implantable electromagnetic middle-ear bone-conduction hearing aid device

US4817607	1989	Magnetic ossicular replacement prosthesis
US4936305	1990	Shielded magnetic assembly for use with a hearing aid
US5163957	1992	Ossicular prosthesis for mounting magnet
US5259032	1993	Contact transducer assembly for hearing devices
US5554096	1996	Implantable electromagnetic hearing transducer
US5772575	1998	Implantable hearing aid
US5941814	1999	Arrangement for adjusting and fixing the relative position of two components of an active or passive hearing implant
US6077215	2000	Method for coupling an electromechanical transducer of an implantable hearing aid or tinnitus masker to a middle ear ossicle
US6682472	1999	Tinnitus rehabilitation device and method

The first simplification of the network, useful for understanding the overall scenario is based on the selection of the patents that can be reached from main paths with only a given number of steps, or based on a threshold in the link count. From the 37 patents in the three disjoint main paths the network reachable in 3 steps is made of 1132 patents, but this reduction is not sufficiently good because it selects not relevant patents. If instead the above selection is reduced by introducing a threshold of 0.0005 in the weight the resulting network of 232 patents has less noise and the resulting layout can be better understood.

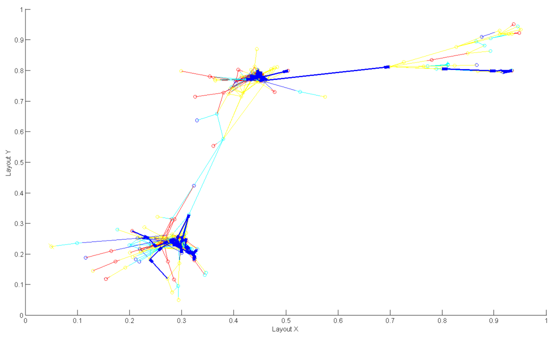

Figure 15 Graph of the reduced network highlighting the three major topics in the patent set, obtained from the three main trajectory and then extended by propagation. Patents and citations are shown as circles colored by distance from the main paths, while the main paths are colored in bold blue.

The second phase of the analysis is related to the identification of the major actors involved in the topic, based on the assignee information. After normalization and removal of assignees with less than 4 patents, the result is of 242 with a mean of patent for each assignee of 6. Examples of top assignees are MedTronic (139), Siemens (42) and Kendall (31).

IV.i Hearing Devices

In this topic there are 5 major assignees in the main trajectory of 15 patents, while there are 75 in the overall sub-network. For each assignee it is possible to compute the evolution of its portfolio along time based on the reciprocal of the geodesic distance of every patent with the respect to the main path. The evolution of the best 5 assignees along time is shown in Figure 2.

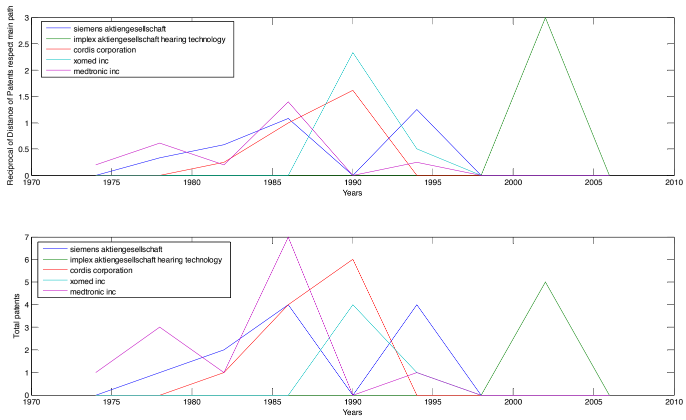

Figure 16 Evolution of the portfolios of the 5 most relevant companies in the topic. The upper plot shows the distance of patents respect the main path, while the lower shows the total number of patents in the topic. Note that the Xomed company has merged into MedTronic on 1999.

For investigating in detail this topic we built a sub-network by propagation up to 5 levels creating a sub-networks of 592 patents. The information from the citation relationships and the company portfolio can been integrated in a synthetic view presented in the form of map (Figure 3). The map presented in Figure 3 has been obtained first by generating a force directed layout of the sub-network and then using the density of patents for representing height and color.

The interactive aspect of the tool used for this work is in the possibility of zooming in and out the map, selecting single patents for obtaining information like title and year and highlighting different trajectories and assignees.

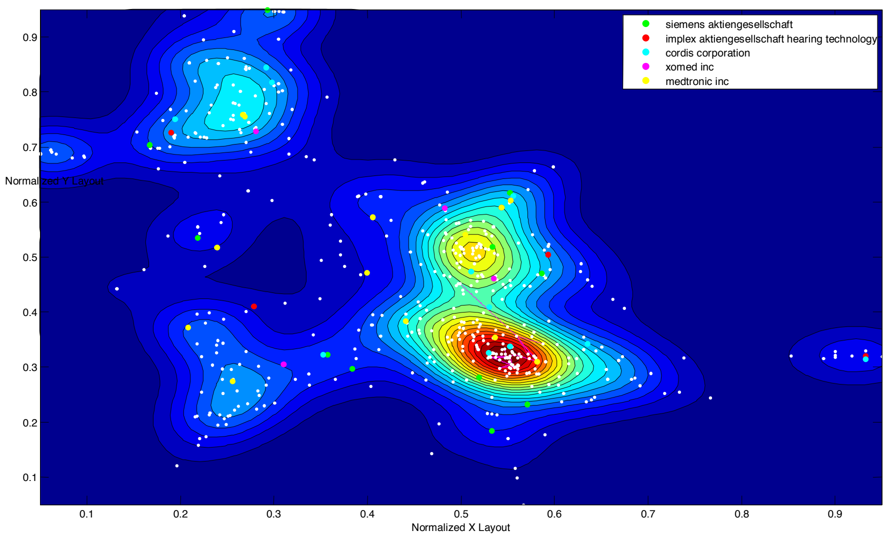

Figure 17 The map of the patent sub-network in the area of hearing rehabilitation devices. The single patents have been placed over the map as white dots for generic patents, or with colored dots for one of the 5 most relevant companies. Finally the main path has been highlighted in purple. The central isle in red is the one associated by the topic of hearing device while the smaller isles are relative to supporting technologies like electronic interfaces, implanting or

measurements. For example the rightmost island is related to implants and this is the reason why Cordis has some relevance there.

IV.ii Knee Braces

The main trajectory of this topic is made of 13 patents and the propagated network at three levels contains about 800 patents. When looking at the map associated to this network it is possible to identify the core of the main path and then some related topics about exercising device and posture control (Figure 4).

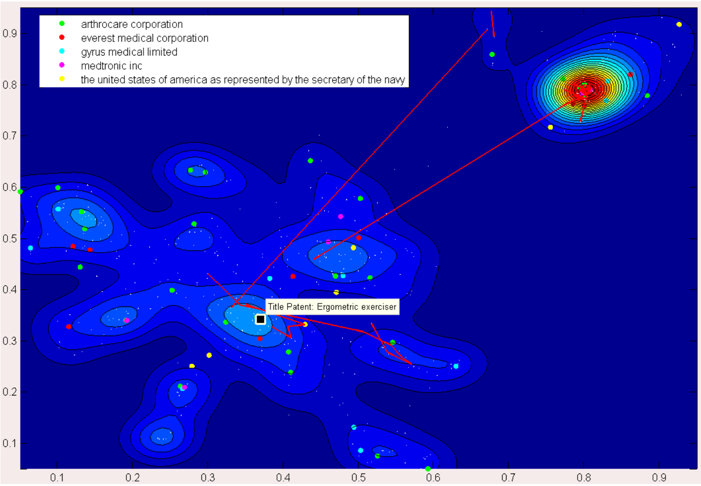

Figure 18 Map of the sub-network generated by the topic of the second trajectory. This figure has been taken from the interactive tool while obtaining information about a specific patent. In the upper left part the core of the topic is present while in the lower left there are associated technologies

In terms of evolution along time this topic has received more attention than the other in recent years, and in particular there is a strong set of patents near the main trajectory, associated to new companies as Arthrocare (1998) and Gyrus Medical (Figure 5).

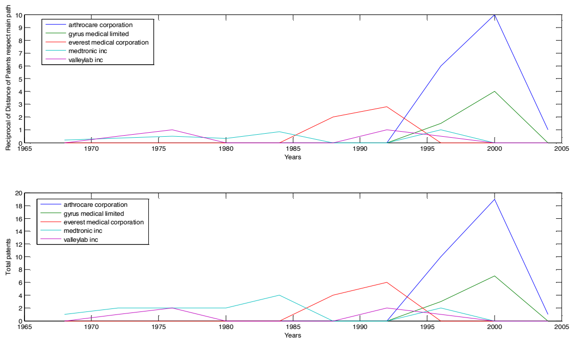

Figure 19 Evolution map of major players in the area of knee braces, with the same type of plot as Figure 2. This plot shows how two recent companies have a strong role in the trajectory associated to this topic.

V- Main conclusions

This work introduces a methodology for the visualization and exploration of a patent landscape based on the analysis of citation networks of patents and the extraction of the major trajectories. The integration of term-based analysis of patents' texts will provide an approach for the visualization of the surface in alternative to the layout based on the citation graph.

The main limitation of this work is related to the fact that patent analysis covers only the technology patented discarding the ones that are not protected or the minor inventions. In the second place, due to procedural reasons there is a gap based on the Priority Date along with the Publication Date: the former represents the first date of filing of each patent, while the latter is the first public disclosure of the patent itself. This causes a delay in the identification of a possible interested new technology.

References

Abernathy W., Utterback J. (1978) Patterns of industrial innovation. Technology Review June, 40-47

Albert M., Avery D., Narin F., McAllister P. (1991) Direct validation of citation counts as indicators of industrially important patents. Research policy 20, 251-259

Anderson P., Tushman M.L., (1990) Technological Discontinuities and Dominant Designs: A Cyclical Model of Technological Change. Administrative Science Quarterly, 35

Boyack K., Wylie B. & Davidson G. (2002), Domain visualization using VxInsight for science and technology management. Journal of the American Society for Information Science and Technology, 53, 764-774

Dosi, G. (1982) Technological Paradigms and Technological Trajectories: A Suggested Interpretation of the Determinants and Directions of Technical Change Research Policy, 11, pp. 147-162

Eldredge N., Gould S. (1972) Punctuated equilibria: an alternative to phyletic gradualism. In: Schopf, T. (Ed), Models in Paleobiology, Freeman, San Francisco, CA

Fleming L., Sorenson O. (2001) Technology as a complex adaptive system: evidence from patent data. Research Policy 30, 1019-1039

Fleming L., Sorenson O. (2004) Science as a map in technology search. Strategic Management Journal, Volume 25, Issue 8-9, 909-928

Grupp, H. & Schmoch, U. (1992) Perceptions of scientification of innovation as measured by referencing between patents and papers: dynamics in science-based fields of technology Dynamics of Science-Based Innovation, Springer Verlag

Hall B., Jaffe A., Trajtenberg M., (2000) Market Value and Patent Citations, NBER working paper series, ISSU 7741

Helo, P. (2003), Technology trajectories in mobile telecommunications: analysis of structure and speed of development International Journal of Mobile Communications, Inderscience Publishers Ltd, 1, pp. 233-246

Huang Z., Chen H., Yip A., Ng G., Guo F., Chen Z.K., Roco M.C, (2003), Longitudinal Patent Analysis for Nanoscale Science and Engineering: Country, Institution and Technology Field. Journal of Nanoparticle Research, 5, 3, 333-363

Hummond, N. P. & Doreain, P. (1989), Connectivity in a citation network: the development of DNA theory, Social Network 11, 39–63

Kauffman S. (1993) The origins of Order: self-organisation and selection in evolution. Oxford University Press, New York

Martinelli, A. (2008) Technological trajectories and Industry evolution: The case of the Telecom Switching Industry. DRUID-DIME Accademy PhD Conference

Rinne M., (2004) Technology roadmaps: Infrastructure for innovation, Technological Forecasting & Social Change, Elsevier, 2004, 71, 67-80

Rosenberg N. (1982) Inside the Black Box: Technology and Economic Change. Cambridge University Press, New York

Sani E., Frisoli A., Bergamasco M. (2007), Patent based analysis of innovative rehabilitation technologies, Virtual Rehabilitation, IEEE, 27-29 Sept. 96-101

Tory M., Sprague D., Wu, F., So, W. & Munzner T. (2007), Spatialization Design: Comparing Points and Landscapes. IEEE TRANSACTIONS ON VISUALIZATION AND COMPUTER GRAPHICS, IEEE Computer Society, 2007, 1262-1269

Verspagen, B. (2007) Mapping Technological Trajectories as Patent Citation Networks. A Study on the History of Fuel Cell Research, Advances in Complex Systems, 10, pp. 93-115