Università
Ca' Foscari
Venezia

Scuola Dottorale di Ateneo
Graduate School

Dottorato di ricerca
in Economia Aziendale - Business
Ciclo 24
Anno di discussione 2013

Wiring knowledge
Metaphors, boundaries and scientific performance

SETTORE SCIENTIFICO DISCIPLINARE DI AFFERENZA: SECS-P/08 e SECS-P/10
Tesi di Dottorato di Claudio Biscaro, matricola 955666

Coordinatore del Dottorato

Prof. Anna Comacchio

Tutore del Dottorando

Prof. Massimo Warglien

Co-tutore del Dottorando

Prof. Anna Comacchio

TABLE OF CONTENTS

ACKNOWLEDGMENTS

Sometimes it's difficult to understand what generates a good idea. Especially in these years of my doctoral journey, I realized that outcomes of meetings and conversations cannot be forecasted. Intuitions come unexpectedly. Often without an apparent reason. Thereby, I cannot deny the importance of the social interactions. Through these years, I met an incredible number of great individuals: advisors, colleagues, friends and family played different, yet meaningful, parts in this successful and very rewarding time.

If the doctorate is a learning process, the story of my development will be primarily a story of trust and of a mounting effort to meet mentors' tacit expectations. I am extremely grateful to my main advisor Massimo Warglien who introduced me to strands of cognitive science literature that guided me throughout the research. He allowed me to freely explore unconventional research paths whereby I discovered a growing passion for computer programming. Working with him unconsciously made me increase my standards. I specifically acknowledge the role of two probably underestimated courses: Computational modeling and the crash one in Processing. I also owe much gratitude to Anna Comacchio who closely followed my work in the last 2 years. Discussions with her were simply illuminating and the gratitude and joy for developing a paper together will be borne for years to come. Then comes Richard Whittington, who mentored me in the Oxonian time. He friendly (and almost weekly) revised drafts of my papers, and taught me the arts of an accurate *labor limae*. He and Victor Seidel guided the maturation of a young doctoral student. Last, I thank Carlo Giupponi, who also almost blindly trusted me and allowed me to pursue my goals in my scholarly networks.

Then my friends. Francesco Reato has a special mention. He is a genius with broad interests who helped me with smart coding solutions and to develop a very useful piece of software. Without him, a part of this thesis would have been more bumpy. I'd like to thank also Vahid and Fish with whom I filled much of my spare time with smart conversations and silly moments. Sonia, for the fruitful collaboration in the first year. Siddharth Arora and Juan Pancho Martinez were also truly good friends and kept active my mind in the damp weather. Lili and Laura for beautiful and long chats over coffees.

My readers. Sometimes I wonder why they did, because reading something you barely understand can be quite boring, but I thank Ben, Sasha, Trudi, Sara, Davide, and Francesco for reading, commenting, and editing my drafts.

Ultimately my family. My dad Giovanni, my mum Dila, and my two brothers Giulio and Fabio for their support to make it till the checkered flag. Mum and dad, in the continuous sacrifice for their sons, embodied the principle that true happiness comes from giving. I've still to learn from you.

Grazie.

WHY KNOWLEDGE COMBINATION MATTERS?

Combining knowledge is key for generating new and better ideas. It has been shown in diverse contexts such as firms and scientific labs, by analyzing groups, teams, and also individuals. Knowledge combination is yet so important that has been studied in diverse research fields: management, psychology, cognitive science… even physics.

To study knowledge combination and innovation, there is probably no better field than science. The scientific world has two properties that make it a unique setting: is continuously chasing new ideas, new inventions, new discoveries that become old as soon as they are published. Moreover scientists are extremely specialized individuals. They have been trained for years in their own sub-discipline and their knowledge is so deep that requires idiosyncratic jargons and constructs and they handle tools that can hardly be applied in other disciplines. And if specialization comes with its efficiency advantages, costs of bridging the disciplinary boundaries are around the corner. Thereby, scientists, on the one hand, continuously try to generate new knowledge and they can do it both on their disciplinary domain, or by collaborating with scientists with other background. In this case, they must reduce the costs that come with failures, inefficient working, and learning costs to understand each other.

In this thesis, the two common threads tying the three chapters are knowledge combination and the scientific setting. In the first chapter, we look at the process of knowledge combination and at the role of language, namely metaphors, in facilitating three differently specialized scientists to guide disciplinary and multidisciplinary research. In the second chapter, we illustrate a method to show the concepts and tools that are at the boundary between disciplines and we make use of scientific abstracts used in a multidisciplinary collaboration. In the third chapter, we look at the articles published within a scientific field and we show that the scientific impact of an article are positively influenced by the number of previous co-authorships that authors have established and by the effort of combining fragmented knowledge.

1. WIRING KNOWLEDGE DOMAINS. METAPHORS AND KNOWLEDGE COMBINATION IN A MULTIDISCIPLINARY FIELD

Claudio Biscaro — *Ca Foscari University of Venice*

Anna Comacchio — *Ca Foscari University of Venice*

Massimo Warglien — *Ca Foscari University of Venice*

ABSTRACT

In the process of combining distant domains of knowledge, metaphors play a privileged role in constructing a shared understanding and in coordinating multiple actors with different background, language and practices. Yet they are still relatively under-investigated, in particular as regards their dynamic interplay with individual cognition and action in the knowledge creation process. Through the case study of a neuroscience research project over a eight-year period, we reconstruct the role that metaphors play in defining conceptually the object of research, interfacing and coordinating different bodies of knowledge, and informing actual practices of laboratory experimentation and technology development. We show how metaphors develop and change over the different phases of the project, responding to the new puzzles they contribute to create and to the changing composition of the network of actors involved. We offer some insight on the emergence of such metaphors and their dynamics in processes of knowledge combination.

keywords: knowledge combination, metaphor, metaphor dynamics, similarity recognition

INTRODUCTION

> "To metaphorize well implies an intuitive perception of the similarity in dissimilar" Aristotle.

Since Adam Smith (1776), it has been largely recognized that innovation is the outcome of a process of recombining different domains of knowledge. In the last couple of decades, the reincarnation of this long-standing view in the debate on innovation and knowledge combination (Kogut, & Zander, 1992; Hall, 2000) has prevailingly focused on the features of knowledge domains recombined in the innovation process, such as their degree of tacitness (Nonaka, & Takeuchi, 1995), modularity (Baldwin, & Clark 2000), technological distance (Nooteboom, Van Haverbeke, Duysters, Gilsing, & van den Oord 2007), or on firms' characteristics facilitating knowledge transfer and combination, such as absorptive capacity (Cohen, & Levinthal, 1990; Zahra, & George 2002). Comparatively less attention has been devoted to the complex process through which such combination is achieved (Faraj, & Sproull, 2000; Leonardi, 2011). Moreover, research on innovation management has mostly focused on knowledge combination as a transfer process, comparatively neglecting the phenomenon of the modification of knowledge through recombination and the creation of a genuinely new knowledge domain.

In this paper, we join the relatively less developed thread of research focusing on the process of new knowledge generation by the combination of heterogeneous sources of knowledge. Previous research on the combination process has focused on issues such as knowledge transformation[1] (Carlile, 2002), knowledge brokerage (Hargadon, 2002; Hargadon, & Bechky, 2006; Hargadon, & Sutton, 1997; Sutton, & Hargadon, 1996) or the role of objects in coordinating and solving boundary spanning problems (Bechky, 2003; Carlile, 2002; 2004; Hsiao, Tsai, & Lee, 2011; Levina, & Vaast, 2005).

[1] As Carlile (2002: 445) defines, "transforming knowledge refers to a process of altering current knowledge creating new knowledge, and validating it within each function and collectively across functions." This process is required when there is difference across knowledge bases, dependence as well as novelty.

This body of literature mostly highlights the social interplay needed to transform knowledge. We focus here instead on the role that language and cognition play in the process of coordinating actors with large differences in their disciplinary knowledge background and language, loci of work, practices and tools in their effort to generate genuinely new, cross-disciplinary knowledge. In particular, we focus on metaphors, a language trope whose centrality in our conceptual system has been increasingly recognized (Lakoff, & Johnson, 1979; Ortony, 1993; Gärdenfors, 2000; Gibbs, 2008). Metaphors are language expressions that open a window on the nature of human mental representations and thinking; they live in linguistic communication, but reveal important aspects of our inner mental processes. They are thus ideal lenses over the process by which actors with different systems of concepts and categories grapple to achieve a common understanding of something novel to all of them.

Metaphors are increasingly calling the attention of management scholars due to their creative role for sensemaking and directing collective change (Hill & Levenhagen 1995; Cornelissen & Clarke 2010), and as they aid actors to build new knowledge from prior experience (Dunbar 1997; 1999). Notwithstanding recent contributions on grounding understanding (Clark, 1996) in situations in which distance between domains is high and communication is unproductive (Bechky, 2003), little is known about how metaphors work as a mechanism of communication and knowledge creation across actors with diverse disciplinary backgrounds. Therefore the aim of this paper is to analyze metaphors in processes of knowledge generation across distant disciplines.

We focus on the dynamics of metaphors addressing the question of how metaphors work throughout the combination process. Our aim is to extend research on how metaphors emerge, and how they frame activities and change by bridging multiple actors and disciplines over time. First we analyze how metaphors emerge as heuristic instruments that not just cast light on novel features in a specific discipline (Black, 1979; Cornelissen, 2004), but help to mould a new cognitive domain bridging multiple input knowledge spaces (Fouconnier and Turner 1998). Secondly, we study how metaphors interact with actions (Dunbar 1997; 1999) and evolve over time in a building process of a multidisciplinary knowledge domain.

Through the analysis of the development of a nanotechnology research project over a time period of eight years, we reconstruct the role that metaphors play in defining conceptually the object of research at the project inception, interfacing and coordinating different bodies of knowledge. Moreover we focus on the metaphors dynamics. On the one hand we show metaphors' impact on researchers actions: looking at their facilitating role in the identification of intermediate adjustments and investigating how they inform actual practices of laboratory experimentation and technology development. On the other hand, we show how metaphors develop and change over the different phases of the project, responding to the new puzzles they contribute to create and to the changing composition of the network of actors involved.

In the next sections of the paper, we first develop the theoretical framework. In the method section we describe our exploratory and longitudinal fieldwork. In the following section we reconstruct the role that metaphors played in the effort of knowledge combination in the main three experiments of the project under investigation. Then we discuss the heuristic power of metaphor and cast light on the impact metaphors have on actions and the subsequent impact actions have on the life and use of metaphors. Finally, we draw conclusions and theoretical implications.

THEORETICAL BACKGROUND: A SHORT DETOUR ON METAPHORS AND KNOWLEDGE COMBINATION

Generating knowledge through combination

A powerful mechanism to generate new knowledge that sometimes translates into breakthrough innovations has been identified in the combination of distant knowledge domains (Hargadon, & Bechky, 2006; Hargadon, & Sutton, 1997; Nooteboom *et al*. 2007). Giving salience to this idea, it has been recognized that some firms exploit the opportunity to broker between communities of practice who are detached from each, because it allows them to use insights and apply methods from one domain to solve problems of the other domain in non-ordinary fashion. Yet, this process has focused scholarly attention mainly on the structural advantage of the brokering position of actors (Hargadon, & Sutton, 1997; Hargadon, 2002) and on the transfer of knowledge from a source to a recipient (Argote, & Ingram, 2000; Jensen, & Szulanski 2007).

Indeed, generating knowledge through a combination of knowledge domains is not only a process of transforming and translating knowledge for a new community who does not know it and who must recognize its value to be able to re-use or adapt the extant knowledge through an idiosyncratic process of transformation (Carlile, 2002; 2004). In short, it is not simply a transfer problem. For instance, in a case of collaborative endeavours whereby heterogeneous groups contribute to create something new, the process of transformation is mutual and the knowledge transfer may be limited to the essential, because it is costly and it could be also inopportune. In other words, a team of specialized actors needs to coordinate around a common understanding of artefacts, a common goal, but at the same time actors cannot swap too much of their specialized knowledge, as it requires extensive training to be accumulated and used.

Language, then, is a necessary element to study in order to keep track of the whole process of knowledge combination, because it is the main driver of information and communications between actors and it displays how cognitive domains adjust, influence and are modified by common goals. We thus analyse the linguistic mechanism of metaphors that have the power to draw meanings from a familiar context to understand what is not familiar. In addition, metaphors are not just mere knowledge and meaning transfer tools, as they embed the ability to mix elements from different cognitive domains in order to generate entirely new meaning.

A short detour on metaphors in knowledge combination

Traditionally, metaphor is defined as “a figure of speech in which a word or phrase is applied to an object or action to which it is not literally applicable” (Oxford English Dictionary). It is often associated to poetical or rhetorical use of words: the figurative use is opposed to the literal one. However, after being relegated for long time to the peripheral domain of language artifacts and rhetoric, in the last decades metaphor has come to occupy the central stage of our understanding of human thought and action. In their influential book on "Metaphors we live by", Lakoff and Johnson (1979: 4) have stated that "our ordinary conceptual system, in terms of which we both think and act, is fundamentally metaphorical in nature."

Cognitively speaking, metaphors provide an understanding of things and events in terms of other things or events. Metaphors can structure much of our daily experience (as in the conception of time as space), but they are also fundamental in dealing with previously unexperienced situations, by projecting what we already know about a domain onto the new domain and thus shaping our understanding of it. Thus, metaphors help us to give meaning to new experiences and objects, and they provide systematic guidance to generate inferences and direct action in relatively unknown contexts. They may also offer a guide to imagine new things, by providing a way to structure relationships among elements according to the structure of the original domain of the metaphor - metaphors can be "generative" (Schoen, 1993). For example seeing a DNA strand as "two-legged" helps to imagine a piece of DNA that "walks" along another strip of DNA (Shin, & Pierce, 2004).

Of course, since metaphors are not identities, they are selective: some elements get hidden while others are emphasized. A mechanistic metaphor of organizations (Morgan, 1986) fatally hides the political side of organizing .

Metaphors, are not just conceptual, they also affect actions. We "live by" metaphors because the way we act and react is structured by the expectations and even the values which are carried by a metaphor. Conceiving a market competitor as "enemy", will favor aggressive marketing behavior (Rindova, Becerra, & Contardo, 2004).

Finally, metaphors are fundamental to share concepts and coordinate action with others. Metaphors are powerful vehicles to convey and translate to others meanings that could hardly be expressed literally (as for emotions), or that could be not understood if expressed literally (as for specialized, idiosyncratic knowledge). Metaphors have been identified as cause and driver of collectively coordinated efforts towards a goal (Dunbar, 1997; 1999) and for their power to attribute meaning to the world, recruit external aid and provide the legitimacy and familiarity to new scenarios (Cornelissen, & Clarke, 2010; Cornelissen, Holt, & Zundel 2011; Hill, & Levenhagen, 1995).

Metaphors are also an effective mechanism to generate new knowledge within disciplinary groups. As a recent study suggests, they are used abundantly in research projects in laboratory meetings to discuss issues and results of research, they provide hints and methodology to solve scientific problems, thus enhancing disciplinary knowledge. Most used by researchers in knowledge generation are those metaphors that bring in knowledge from very close domains, while those, which rely on distant domains, have an explanatory function (Dunbar 1997; 1999). Yet the role of metaphors in multidisciplinary contexts is relatively neglected by the literature on innovation.

In multidisciplinary processes of new knowledge combination, novelty recognition is just one problem to be solved. A more complex endeavour of knowledge building through combination has to be favoured. While the role of metaphors in novelty representation within a single discipline has been analyzed (Dunbar 1997; 1999), less is understood when novelty is at the intersection of different disciplines. Thus, we want to contribute to the extant literature analyzing how metaphors emerge in research efforts at the intersection of disciplines.

Moreover being used as cognitive tools, metaphors might be challenged or even disconfirmed while the research of novelty proceeds. Indeed these dynamics are still under investigated. Thus we want to study the metaphors dynamics, focusing on the interplay between metaphors, interpretations and actions along a process of knowledge combination.

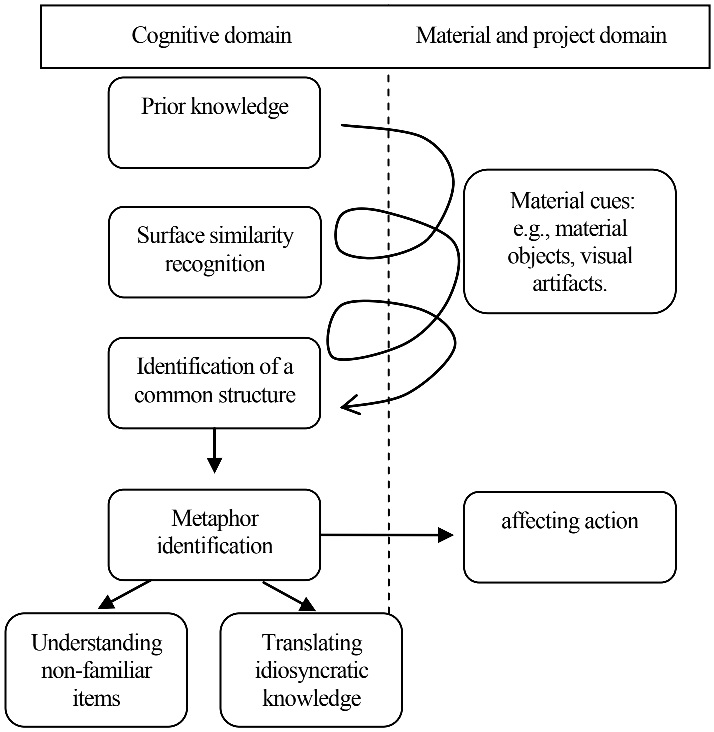
How metaphors bridge cognitive domains

According to this aim, our approach is to adopt a cognitive view of metaphors that defines their structure in terms of mappings between a "source" concept and a "target" one (Lakoff, 1993). For example, the metaphor "Love is a journey" maps the source domain of journey to the target one of love. Elements in the source are mapped onto corresponding ones in the target (e.g. travelers -> lovers; the vehicle-> the love relation). The mapping goes beyond correspondences between elements. It transfers the knowledge we have about the source domain onto the target domain: a process that helps to reason about the target domain and generate new meaning (Grady, Oakley, & Coulson 1999) .

The transfer of knowledge made possible by the source-target mapping relies crucially on the perception of a similarity relationship between source and target (cf. Gregoire, & Shepherd, 2012). It is common to make a distinction between two types of similarity mapping: *surface* similarity, and *structural* similarity. Surface similarity refers to the resemblance across domains of basic attributes such as the color, shape or qualities (Markman, & Gentner 1997), whereas structural similarity (Gentner, 1983) refers to similarity only in the relational structure of the domains, hence in a predicative relationship that links at least two attributes in each domain. The perception and identification of the latter type are cognitively demanding when surface similarities are missing, scarce, or not transparent and cues to interpret different contexts are insufficient (Catrambone & Holyoak, 1989; Gentner 1989; Keane, Ledgeway, & Duff 1994). The distinction is important, because surface similarities can be misleading in the way we transfer knowledge across domains (Gilovich, 1981).

In collaborative endeavors, metaphors are used for cognitive and practical purposes and their role of guidance spans between the cognitive and the material domain, relying on a necessary perception of similarity that is influenced by actors' prior knowledge and material and immaterial cues available, e.g., narratives, objects, images. We will focus only on the material ones, as product of the scientific research. In Figure 1, metaphors' emergence and its spanning capability are summed up.

Figure 1. Contextual emergence of metaphor. Union of cognitive and material domains.



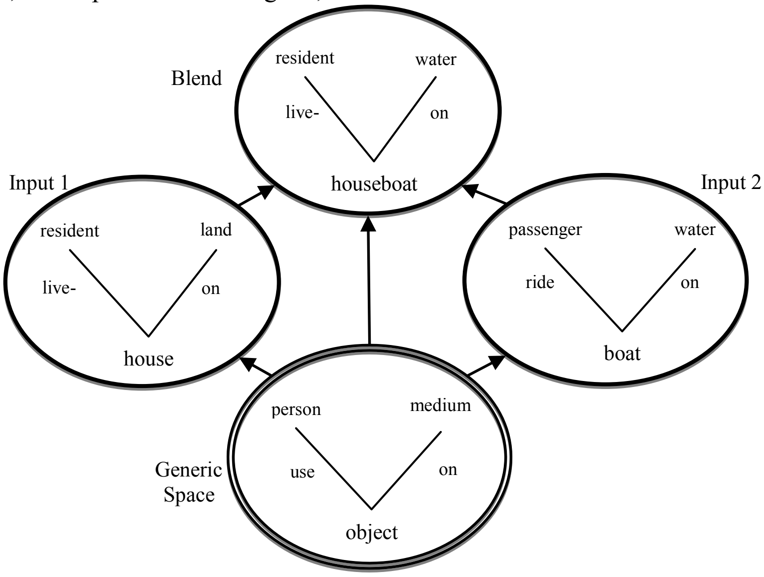
In generative processes stemmed from the combination of different strands of knowledge, the role played by the perception of similarity has been controversially explained by management scholars who on the one hand place it as a pivot for bridging different cognitive domains, and on the other also recognize that similarity constrains creativity. Indeed, between-domain similarities and dissimilarities establishes the boundaries conditions of the creative power of metaphors. For Oswick and colleagues, similarity recognition is a constraint for creativity, suggesting that the differences between domains represent the fertile ground from where to bring new knowledge and insights, hence other tropes that rely less on similar attributes or relations, such as the irony, would be more productive (Oswick, Keenoy, & Grant, 2002; Oswick, Fleming, & Hanlon, 2011). Instead, Cornelissen and his colleagues stress that is the metaphoric conception that moves the recognition of shared qualities across domains (Cornelissen 2004; 2005; 2006; Cornelissen and Clarke 2010) and that creativity instead is due to the distance between domains.

Finally, a cognitive view of the metaphor as a source-target mapping considerably blurs the boundaries between metaphors and other figures of speech such as analogy or simile (Holyoak, & Thagard 1996), reconducing them to “a same basic human ability” to make similarity-based mappings (Fauconnier, & Turner, 2006) - a point many psychologists would challenge. We will not delve here in this controversial subject, as for our goals Lakoff’s (1993) mapping definition will suffice as a first approximation.

Metaphors, blend and generic space

Instead, our study draws on a useful enrichment of the basic mapping structure introduced by Fauconnier and Turner (1998) in their work on conceptual integration, in order to understand the role of metaphors in processes of new knowledge development across distant domains. Fauconnier and Turner stress that the source-target mapping is just a special case of richer structures of conceptual integration, that relies on the combination of multiple sources into an integrative target (or, in their terminology, blend). So, for example, houseboat results from the mapping of two sources (the house and the boat) onto the new object (Goguen, 1999). The conceptual integration extension shows that metaphors are creative to the extent to which combining two semantic domains leads to construct correspondences, which were not there prior to the metaphoric thinking (Cornelissen, 2005). Specifically, this generalization of the source-target mapping will turn to be very helpful in what follows, since, as we shall see in multi-disciplinary contexts, a metaphorical target domain has to be simultaneously an integrative cognitive structure of multiple source domains to effectively mobilize different disciplinary knowledge.

Figure 2. A textual blend, an adaptation from Goguen, 1999.



Furthermore, Fauconnier and Turner have introduced a third and abstract domain (the generic space) that is a sort of common ground for the source domains, providing the basic structure within which the similarity between sources and their mapping to the target can be defined. In the houseboat case, the generic space might be the person-object-medium space of Figure 1 (adapted from Goguen, 1999). The notion of a generic space is important, because it shows that similarity between domains, needed to support their mapping onto the metaphoric target, can only arise within the broader structure of the metaphor itself (a point already clearly made in Lakoff, & Johnson, 1979).

Our intent is to study the dynamics of metaphors in the process of knowledge generation when this involves distant disciplines, we address the question of how metaphors work throughout the knowledge combination process. We aim at extending research on how metaphors emerge, how they frame activities and change, and bridge multiple actors and disciplines over time. We claim that metaphors emerge as a heuristic instrument and they not solely shed light on novel features in a specific discipline (Black, 1979; Cornelissen, 2004), but they help to shape a new cognitive domain by selectively mapping knowledge from multiple, heterogeneous input spaces (Fauconnier and Turner 1998), and contribute to frame the joint action of different actors.

METHODS

Research Setting

The NEUR multidisciplinary research. According to our exploratory aim, that is to understand how metaphors emerge and to analyze their role in novel knowledge creation processes across distant disciplines, we adopted a theory-building qualitative research approach based on the case study of a eight-year neuroscience research project (Eisenhardt, 1989; Strauss and Corbin, 1990).

We chose to study a single case with in depth analysis, because it this research setting offered a unique case of a interdisciplinary radical research over several years with a deep potential for theoretical contributions (Siggelkow 2007) on the role of metaphors in knowledge combination processes and their dynamics.

We investigated a setting in which knowledge creation and a multidisciplinary context were key features. The research setting was the NEUR[2] research project, a eight-year scientific research that started in 2002 with a national grant and two partners and continued with a further project, financed by a leading European Institution in 2006, which involved three more research institutions and a commercial partner. In total six institutions collaborated and they were located in six countries (Italy, Belgium, Switzerland, France, Israel, Germany).

The NEUR research project proved to be a successful under different metrics. In terms of fund raising, scientists were able to obtain two important grants: a National one and a very prestigious and competitive European one. 41 articles have been published in peer-reviewed journals based on the work of the various partners. With respect to the scientific impact, both journals' impact factor and the number of forward citations obtained by the three main joint-experimental articles (well above one hundred for each of them) makes them a successful group. Moreover, this group has established a first theoretical model of the interaction between neurons and carbon nanotubes.

[2] All names have been encrypted to ensure confidentiality.

The NEUR primarily aimed at developing a new generation of implants to repair damaged central nervous systems (CNS) tissues. Specific research goals branched off from the main one. Minor goals were to advance research on biophysical interactions between nanomaterials (carbon nanotubes) and neurons, to exploit nanomaterials as environment to favor damaged CNS tissues regeneration and to fabricate new neural microelectrodes. Indeed, it was a research project at the scientific frontier of distant disciplines such as medicine, biology, engineering, chemistry and physics. The NEUR was also heterogeneous in the research practices of different laboratories and in differing types of experiment: in vivo, in vitro, and in silico (computer simulations).

Whereas usually in multidisciplinary research, "various disciplines address scientific and social challenges independently" (OECD, 2010), NEUR succeeded in integrating and building novel knowledge by a deep effort of interaction across distant domains and different researchers and laboratories. In doing so, scientists achieved significant new understanding and novel results in the emergent research area of the modern nanotechnology application to biological systems.

In our study we focused on the interaction of the three core disciplines and scientists' background with limited shared knowledge and practices, as shown in Table 1. Medicine, specifically the neurophysiology (henceforth NPH) lab was specialized in electrical recordings of the activity of the central nervous system (recording and resistance = 8; explants = 6) and had a solid background in statistics to deal with the high variability common of biological systems (statistics, statistical analysis section = 8). The Chemistry (CH) lab was specialized in handling carbon nanostructures and characterizing them with different properties (carbon structures and functionalization = 8), whereas the Engineering (ENG) lab had expertise in formal modeling of neural behavior (voltage, and presence of equations = 8). The overlapping tools or techniques in the laboratory activities were very few (e.g., voltage, microscopy).

Table 1. Disciplinary techniques, tools and activities.

techniques/tools/activities	NPH	CH	ENG
Voltage	8	1	8
in vitro (or on glass)	8	1	2
statistics (st. significance/ test / st. analysis)	8	0	6
recording (single cell / population)	8	0	2
resistance	8	0	2
statistical analysis section	8	0	1
explants	6	0	0
pharmacology (blocks, inhibitors, facilitators)	5	0	1
inverted microscope	4	0	0
noise	2	0	7
electric spikes / bursts	2	0	2
electron microscope (TEM)	1	3	0
computer simulation & algorithms	1	0	5
correlation	1	0	3
epifluorescence microscopy	1	0	0
infrared microscope	1	0	0
laser scanning microscope	1	0	0
videomicroscope	1	0	0
carbon structures (fullerene/nanotubes)	0	8	0
functionalization	0	8	0
atomic force microscope	0	2	0
magnetic force microscope	0	1	0
scaffold	0	1	0
scanning probe microscope	0	1	0
equations	0	0	8
microscope (not specified)	0	0	1

Techniques/tools/activities have been counted (1) if present, one or more times, in a scientific publication. E.g., correlation has been mentioned one or multiple times in an article of NPH and in three article by the ENG group.

Informants included the three leading scientists of the project, experts of each discipline. They were the key individuals to interview for three reasons: their advanced background in each of the three core disciplines, their pivotal role within the NEUR project – one of the three researchers was the scientific coordinator of the NEUR project – and their leading position in the laboratories in charge of main experiments and research advancements. Furthermore, they generated and developed the idea for the experiments we will discuss in this article.

Key scientists. The chemist (CHt) at the time of the NEUR's start led a laboratory operating in an Italian University. The laboratory was an organized, well-equipped space where, on average, more than 10 researchers worked. The CHt, full professor since 2002, had an excellent international standing. His contribution in advancing the field of nanoscience, internationally recognised, was based on large body of publications on first-tier chemical journals, as shown in Table 2. He was specialised in handling and functionalizing carbon nanomaterials by combining specific groups of atoms with the pristine carbonic structure to change its properties.

The neurophysiologist (NPHt) from 2002 to 2005 worked in a laboratory of a research center part at an advanced school, that offers postgraduate training in Physics, Mathematics and Neurosciences. Such laboratory was located in the same city of the university of the CHt. The NPHt is an Italian physician specialized in electrophysiology, assistant professor in the same university of the CHt since 2002. While she (NPHt) was still appointed at the research center, the NPHt performed a research project on damaged neuronal tissues. In 2005 she successfully applied, as scientific coordinator for a European grant (2006-2009). Funds allowed the start-up of a new laboratory, where she conducted new experiments with two doctoral students. Before the beginning of the NEUR project, the NPHt had several international publications in her specific research field. She was specialized in explants of neuronal and nervous tissues from rats and subsequent electrical recordings.

Table 2. Research Setting

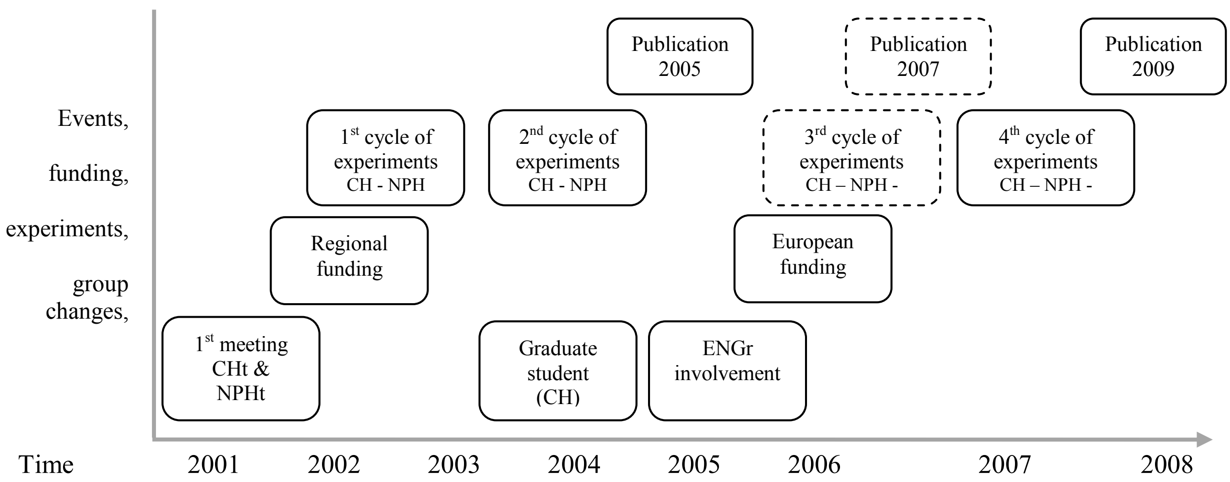
Laboratory - main field	Chemistry (CH) lab	Neurophysiology (NPh) lab	Engineering (ENG) lab
main research area	functionalisation of nanotubes	in-vitro experiment with neurons	computational simulations of neuronal models
Referent	head of the lab	assistant till 2005, head of the lab since then	assistant in Switzerland (2005-2009), head of the lab in Belgium (2008-2009)
Location	Italian University (2002-2009)	Italian Research Centre (2002-2005), Italian University (2005-2009)	Swiss University and Belgian University
Size	10-15 permanent researchers	0-2 researchers till 2005, 2 permanent researchers since 2005	1 researcher in Switzerland, 2 in Belgium
Tenure of referent when he or she joined the project	full professor	assistant professor	assistant professor
Publications of the referent before joining the group*	157	15	15

*Google Scholar ® and curricula vitae have been used as a source of information, data have been screened afterwards

The engineer (ENGr) was a postdoctoral researcher in a laboratory located in Switzerland until 2008. In this lab, experiments on electrophysiology were performed. In 2008, he moved to a Belgian University, where he was tenured as assistant professor and had the supervision of two young researchers in a new laboratory. When the NEUR project started, he was a young researcher in Bioengineering, the field in which he obtained his doctorate. He joined NEUR in 2005. He developed mathematical and computational models of neuronal activity, therefore he was familiar with electrophysiology recordings of population of neurons as he needs to test his models on 'real' data. Before he joined the NEUR project, he had 15 publications.

The time period analyzed starts in 2002 and covers eight years, from the birth of the original research idea, through two granted joint-research projects: a national project from 2002 to 2005 and a European one from 2006 to 2009. We focus on what the project leaders consider in retrospect the three main experiments of the project, an early one in 2003, a second successful one in 2005 that lead to the first joint article, and a third one in 2008, after the research team was enlarged, as shown in Figure 2.

Figure 2. Storyline of the research project.



Data Collection

We gathered data on goals, processes and scientists' interpretations along the entire length of the NEUR project. Data collection involved multiple sources (Eisenhard, 1989; Yin, 1984). Unlike standard data upon which studies on metaphors draw, that almost exclusively rely on textual and visual materials (cf. Gibbs, 2008), we added data sources bringing from the tradition of management case studies. We favored a cognitive rather than sociological reconstruction of facts, as they occurred before we had access to the field.

We gathered data from four sources: (1) interviews with the leading scientists and other team members; (2) visit of laboratories and a direct observation to a half-year meeting; (3) archival data, including scientific publications, images, powerpoint presentations, sketches and other files provided by our informants; (4) press coverage, comprising written and audio interviews in the Italian as well as International press, scientific journals, reviews, blogs and magazines. Two researchers conducted semi-structured and retrospective interviews with the three key scientists, head of laboratories, who initiated the project, see Table 2 for details. A fourth interview was performed with a former graduate student, member of the chemistry laboratory who played a crucial role during the second experiment. For the intent of reconstructing through each interviewee's narrative the representation of the project's turning points, images and metaphors, that guided respondents in dealing with previously not experienced situations, questions were mainly centered on facts and problem-solving activities spanning from the entrance of the interviewee in the scientific venture to the European project conclusion. However, much room was left for interviewees to freely reconstruct the steps of the research path, how they set up the experiments and interacted each other, how they interpreted results, the process of generation of ideas to advance the project and the role of each disciplinary background in a multidisciplinary research. Thus, personal perceptions and understanding of the process, the main steps and research ideas and metaphors emerged. To motivate scientists to give a thorough and accurate account of facts, we ensured confidentiality (Huber, & Power, 1985). After each interview, the three authors confronted their notes on salient parts on the interview; findings, ideas and interpretations over the major results were then triangulated by (Yin, 1984). One researcher directly observed a half-year meeting in which all groups presented their latest results. Herein, notes were taken, the audio recording was then transcribed, and we followed the same process of triangulation. Along with each interviews, a visit at the laboratory and follow-up informal interviews was performed with key scientists and some of their assistants for approximately one hour each.

Among the secondary sources, we collected three interviews to the neurophysiologist and the engineer conducted by journalists during popular radio broadcasts on science. Another interview was part of the online podcast published in the website of Nature. These podcasts were published in 2005 and 2008. The audio recording of all interviews was transcribed verbatim.

We then were given access to textual and visual materials such as images produced by lab microscopes and notes produced by the three key scientists for internal communication purposes. We then gathered documents produced along the NEUR project for external scientific communication purposes, such as conferences, seminars, summer schools, peer-reviewed scientific publications. We also analyzed the final scientific report delivered to the granting institution at the end off NEUR project, together with a master of science thesis and a doctoral one developed on the same topic of the project by junior researchers who were part of the research team.

To strengthen the internal validity of the accounts and to deepen the understanding, we relied only on information overlaps which must come from at least two sources and also must not be disconfirmed in any other (Miller, Cardinal, & Glick, 1997) and we combined retrospective accounts with real-time secondary sources (Leonard-Barton, 1990).

Table 3. Data sources

Data source	Type of data	Analytical use
	Data Sources and Use	
Data source	**Type of data**	**Analytical use**
Interviews (129 pages)	*Semi-structured interviews* (4). One with each of the key scientists and one with a former graduate student in chemistry to discuss the origins of the project, of the research ideas, and the development of the joint project from each laboratory's perspective.	Reconstructing the history of the project. Investigating scientists' interpretation of facts, development of ideas and dynamics between the disciplinary and the cross-disciplinary activity. Triangulating evidence.
	Informal interviews (4) with the three key scientists and with a NPH post-doc to discuss our interpretations, to gain familiarity with their working materials and representations.	Gaining familiarity with each laboratory's materials, techniques, and representations. Supporting our interpretation and triangulate evidence.
Observations	*Laboratory visits* (3). One visit of the lab of each key scientist.	Understanding the dimension of the team and the endowment of tools and resources available to key scientists.
(90 pages)	*Field notes and transcription of half-year meeting*. Verbatim transcription of the words of scientists explaining their laboratory scientific progress. Description of the misunderstanding and the strategies and tools adopted to solve them.	Controlling for the role of key scientists between them and the other partners. Understanding each laboratory's activity and drawing relations between them. Focusing on the visual cues to solve them.
	Primary sources (14 h 30’, 219 pages)	
Archival data	*Scientific articles* (34). Eight of them on the chemical manipulations of carbon structures, eight on the physiology of neurons, eight on computational models of neural behavior (24 in total) published by the key authors before joining the research group. 10 joint publications.	Content analysis to identify techniques, tools, methods in the 24 non-joint articles. In the 10 joint articles, we highlighted the design of the experiment, the role of scientific partners and their relations, the necessary disciplinary advancements, limits of research, and the presence of metaphors (or cues to metaphors) in the text. Reconstructing the history of the project.
(252 pages)	*Scientific theses* (2). A doctoral thesis in chemistry based on the activity within the lab focused on the preparation and functionalization of carbon nanotubes documenting some of the chemistry activities and results between 2006-2009. A Master thesis documenting the activity performed for the first successful experiment published in 2005.	Providing support to data and interpretation. Reconstructing the history of the project.
	Images (391). Visual material (pictures, plots, sketches, and PowerPoint drawings) produced and used for reasoning about and/or sharing ideas with the other research partners. 24 of them were part of the chemistry world, 15 of the neurophysiology, 132 of engineering.	Providing support to data and interpretations. Verifying of the presence of metaphoric reasoning in the sketches. Giving visual evidence to the reconstruction of the history of the project.

(171 slides)	*PowerPoint presentations* (7). 5 presentations of the engineering group, 1 of chemistry, 1 of neurophysiology of their research activity presented at conferences, summer schools and for internal meetings.	Providing support to data and interpretations. Verifying of the presence of metaphoric reasoning in the sketches.
(114 pages)	*Institutional Report* (1) documenting all the activities and results of all scientific partners complying with the application for the grant.	Analyzing the activity of all partners and drawing relations between partners.
Press coverage (25 pages articles)	*Articles in the press* (19). 7 in the International press, 11 in the Italian press. Among these, two interviews to the key scientists were included in scientific journals.	Controlling the adoption of metaphors used in the dissemination.
(18 pages)	*Press interviews* (2). About the dissemination of the research results to a non expert audience.	Controlling the adoption of metaphors used in the dissemination.

Follow-up emails and documents were sent to confirm interpretations and data collected. Moreover one document comprising our main results and interpretation of facts was sent to the NEUR scientific coordinator, discussed with her to gather feedbacks and comments and finally included in the final scientific report that the European project consortium sent to the granting institution.

Furthermore, the engineer enabled us to access his personal archive of images, sketches and presentations used internally at half-year meeting presentations and externally at conferences and seminars. The archive comprised almost 400 images and seven presentations. Over 80 percent of images were generated by the engineer's work. Five presentations were made by the engineer, one by the chemist and the remaining by the neurophysiologist.

Finally documents and scientific publications were analyzed: the text and the images of all ten international publications[3] of the group of scientists – in the year from the 2005 to the 2009 – along with the latest eight scientific articles published by each key scientist (totally 24 papers of about 8 pages each) before joining the common scientific venture. These were used to establish prior knowledge, methodology and practices. In total we analyzed 34 articles. Nevertheless, two out of the 10 publications were finally excluded from the reconstruction of the story, because they were

[3] Nine peer reviewed joint-articles and one single-authored literature review.

reviews of discipline-specific literatures. In their three experimental papers, we studied also the supporting information with extra data and analysis provided in the website of the journal. We also analyzed articles in newspaper and magazines as well as online press of various types, such as blogs, scientific bulletins, online reviews among the others, that contained entire or excerpts of interviews to the key scientists. In total they were 18 articles, and one of these interviews was published in a scientific journal. They covered the period from their first joint publication in 2005 to the last one in 2009, online in late 2008. We analyzed and codified the scientists' statements reported in scientific articles and blogs where they were interviewed, in order to detect how they represented and disseminated their research results and the role likely played by metaphors. In addition, the 114-page final report, dated January 2010, encompassing all results with respect to the deliverables presented to the granting institution. This last document included the work of all six research partners, whereas we focus on the main three, because the role of the others was peripheral according to our research aims, as one partner was the provider of one tool for electrical recording that was not yet adopted in the phases we treat, another performed independent tests on living mice, and the remainder gave external support to the group of chemists based in Italy.

Data Analysis

In our inductive approach, we reconstructed the main facts of three key experiments of the eight year NEUR research under investigation: the early experiment in 2003, the first successful one in 2005 that led to the first joint publication, both funded by the national grant and the last experiment in 2008 carried out as part of the European research project (2006-2009). To identify overlaps and differences across disciplines prior to the research inception, we classified and counted activities, methods and main concepts of the three disciplinary domains by the analysis of the eight articles written by each key scientist before their participation to the NEUR research.

After completing the transcription of each interview, the three authors gathered to discuss their independently constructed viewpoint. A storyline was reconstructed and an interpretation of each experiment was then written in a narrative form with selected quotes of informants. A table summarizing facts with quotes from all sources as well as images to support facts, ideas and interpretations was created to make sense of the story and allow comparisons across given situations in time. Mainly focused on tracing back the causality of events and on deepening the understanding of phenomena, we incorporated both real-time data, such as images, presentations and articles, with retrospective accounts. This enhanced plausibility and coherence of interpretations and allowed us to control for time. Thus, we divided data by year or half-year, when significant events occurred. In order to increase the internal validity, we exchanged emails and documents with the scientists to discuss and confirm our interpretation of facts.

Being six the partners of the NEUR project, we used the Institutional Report documenting all the phases of the research activities made by each partner to draw relations between them on the bases of their research collaborations. We then isolated the core experiments, the main research results of the NEUR group, and the research partners involved.

At this stage, we identified three similarly developed experimental activities that were necessary for the creation of a conceptual framework (Eisenhardt, 1989) on which our narration of facts and the development of our theoretical findings were grounded.

We codified the quotes by trying to maintain the faithfulness with our informants' account and at the same time being abstract and consistent with the literature. Thereby we generated first- and second-order categories (Corley & Gioia, 2004; Stigliani & Ravasi, 2012) whereby the first-order codes are words extracted by our data and second-order codes are abstract constructs well grounded in the literature we use.

After an iterative process of analysis, data on experiments and NEUR researchers interpretations were cross-checked several times. In order to give a clear account that separates facts from interpretations, we adopted a within-case analysis (Miles, & Huberman, 1984), and more specifically a two-order concepts of narration (Van Maanen, 1979) .

In the following sections each key experiment is reconstructed to understand the role played by metaphors. In each main phase we analyze in depth the emergence of a metaphor, how it guides research ideas, the experimental setting, and defines the constraints to which each specific disciplinary group adapts. Moreover we study how new metaphors are developed and respond to new problems or advancements and to the changing composition of the network of actors involved.

FINDINGS

The research activity of the scientists involved in the NEUR project was a continuous interaction between disciplinary backgrounds grappling with the difficulty of understanding each other domain requirements and materials. During the course of the project, scientists made use of microscopy images, reading materials and give joint lectures with the intent to reach understanding and develop a common research path. They used several metaphors that allowed them to facilitate the design of joint experiments and guide their research activity.

By providing an overview of quotes in Table 4, we capture the key cognitive steps of the NEUR project over time.

Table 4 – Data supporting cognitive processes

Data Supporting Cognitive Processes

Second-Order Codes	First-Order Codes	Representative Quotes
Metaphors	not talking anymore	tiny electric wires connecting two neurons that communicate through electric signals. (CHt)
		Carbon nanotubes can be either conducting or semiconducting, in principle they could be used as assistive devices to functionally and structurally re-connect neurons that do not talk to each other anymore. (CHt)
	tiny electric wires	I think that it was easier for them to understand, because of the assumption that nanotubes are like tiny electric wires connecting two neurons. (CHt)
		This planar supports coated with a dispersion of electrically conductive small wires, infinitesimal needles, allowed us to study …. the interface between an artificial device and … the neuronal tissue. (ENGr)

	scaffold	"CNTs represent a scaffold composed of small fibers or tubes that have diameters similar to those of neural processes such as dendrites" (NanoLett 2005 , p.1107)
		Morphological properties of nanotubes' structure, at the electronic microscopy images, shown a very rough surface, fractal-like, with similarities to the neuronal extracellular matrix. (ENGr)
	percolation	"3-D Morphology suggests long-range electrical connectivity through the idea of percolation and conduction" (Nature 2009, p: 126)
	wires sitting on wires	In this meshwork of carbon nanotubes, the idea of the electric wire is a bidimentional simplification that comes from the assumption that a neurone, whose profile is symmetric, sitting on nanotubes turns into an electric wire sitting on other electric wires. (ENGr)
Similarity recognition (surface similarity)	looking alike	Watching the image, we looked at each other and said "these two look incredibly alike" (NPHt)
		Nanotubes look like a matrix protein using the same microscopy (NPHt)
		These nanotubes are nothing else that ... microscopic graphene sheets coiled to form cylinders which geometrically look extremely similar to structures that have the morphological characteristics of cells that are the high majority of the central nervous system (ENGr)
		The idea of putting together carbon nanotubes and neurons comes first of all because of their structural similarities. (superficial similarities in our context) (CHt)
Similarity recognition (structural similarity)	electric signals	Considered that those tubes have an incredible conductivity and neuronal signals ultimately are electric signals…. these structures … could interact (NPHt)
		At least at the beginning, I think that it was easier for them to understand, because of the assumption that nanotubes are like tiny electric wires connecting two neurons that communicate through electric signals. (CHt)
		This makes me see the same microscopy images in a different way: nanotubes touch each other. Thereby I can think of an electrical circuit between any two points in the net (ENGr)
Cognitive puzzles of generating a multidisciplinary knowledge domain.	finding out a technique	To carry out this experiment we had to find out a technique to paste nanotubes on the surface that isn't easy, but we found it. (CHt)
		The control cannot be done (anymore) on the peptide, but it must be performed on glass…. This was our difficulty…. There are series of procedures to minimize neuronal lesions in order to explant them in the best possible condition. (NPHt)
		We had previously seen (in 2002) that commercial nanotubes are very very dirty, so we could functionalize them. We suspended them, then we solved them to clean them from impurities… It came out a cleaner product so we published the results and we patented it as well. (CHt)
	trying to understand each other	At the beginning, the NPHt and I tried to understand how things worked. I think that the use of metaphors was rather common. (CHt)

		She explains it in her words. She hasn't developed yet a fully interdisciplinary language, in the sense that she can explain some phenomena, but with a very specialist jargon, thereby most of the times it's difficult to understand. (CHt)
		Initially, we only observed this phenomenon, the nanotubes favoured neuronal communication. We imagined it had to deal with nanotubes' electrical conduction, but schematize the neuron-nanotube interface in an electric circuit rationalized the interpretation very much. (CHt)
		Reaching the interaction entails that we understand each other. It took a lot of time before I understood what were the chemical what was the chemistry issue and the property of these materials. (NPHt)
	needing a theoretical model	The ENGr joins after we realized that the effect on the neuronal activity was caused by its interaction with nanotubes. (NPHt)
		Look, I have this problem, I need to understand, I feel that we need a theoretical model to clearly interpret the interaction. With these tools I cannot do more. (NPHt)
		We passed from a phenomenological experience to something more rational. (CHt)
		The ENGr made questions the NPHt never made, for example "how do nanotubes conduct electricity?". Very often they were questions I could not answer. (CHt)

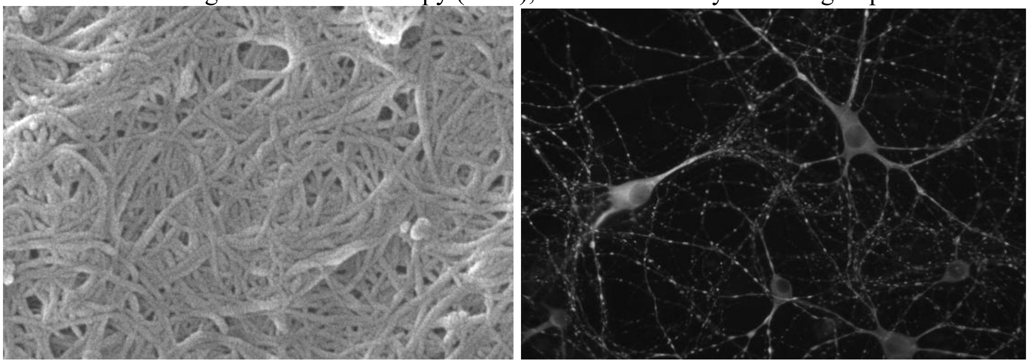
At the NEUR research inception: images and surface similarities

In 2002 the two leading researchers that generated the research idea (CHt and NPHt) were working separately in different institutions and locations and they did not know each other. However an incidental exchange of images was the trigger for their decision to develop a joint experiment.

At that time the CHt was asked to attach a protein on carbon nanotubes to favor the interaction with neurons in order to start experimenting the combination of the two materials. As shown in Table 1, the CH lab's expertise was in manipulating nanotubes, whereas they never handled neuronal tissues, therefore CHt contacted NPHt's supervisor who worked in an important research institute. The researcher suggested NPHt's (his post-doctoral student) to start a conversation with CHt. NPHt was, in fact, a young scientist whose activity was dedicated to understand the electrical behavior of nervous tissues. She dealt with tissues explanted from mice and was familiar with electrical recording tools, pharmaceutical drugs to control neuronal electrical activity and microscopic analysis of living cells.

The images exchanged by the two scientists were two microscopy pictures of carbon nanotubes and neurons, respectively taken by the CH and the NPH laboratories. The casual comparison of the two pictures showed an unexpected similarity between the two objects.

Figure 3. A batch of carbon nanotubes (left) and hippocampal neurons under immunofluorescence (right) at the Scanning Electron Microscopy (SEM), with the courtesy of NPH group.



> "Watching the images, we looked at each other (the NPHt and the CHt) and said these two (nanotubes and neurons) look incredibly alike"(NPHt).

By observing images they noticed the *surface* similarity between the chemical compound and organic material, as seen on the electronic microscope, see Figure 3, and therefore they decided that it made sense try to combine the two materials and start a first joint experiment.

> "The idea of putting together carbon nanotubes and neurones came first of all because of their ... similarities… [n]eurite elongations are reminiscent of the cylindrical shape of carbon nanotubes" (CHt – Pr2Ch05[4]).

The early experiment of 2003 and the *electrical wire* metaphor (metaphor 1)

They initially decided to perform a joint experiment, but it took one year and a half to design it. The two fields were distant and it was the first time that each scientist had to consider the other discipline in one of his/her own experiments. The trivial superficial differences to a non-expert eye between the two representations at the nanoscale were quickly foregone by the CHt and the NPHt who were, instead, attracted by the morphological or superficial–similarities of the two materials, see 'similarity recognition' in Table 4. More specifically the roughness and the branching, see

[4] Secondary source interviews' code.

Figure 2, and by a structural similarity (Gentner 1983), that is the idea that nanotubes, as well as neuronal branching, could transport electric signals. However they knew little about the other scientist's materials properties and behavior, thus notwithstanding the inspiring role of the metaphor of the electric wire, they had to overcome several challenges to design the experiment.

Their effort in identifying possible bridges were focused on reducing the toxicity of the carbon structure and, on the other disciplinary hand, on the way to bridge a broken neuronal tissue with carbon nanotubes. During this period, the NPHt and the CHt worked separately, but sharing reviews, articles and books in order to understand specific features of each other's material domains useful to set up the experiment. To summarize, in the CH lab, it is known how nanotubes might be manipulated in order to be pure and biocompatible, but there was no knowledge about neuronal behavior, properties and requirements in order to make neurons survive in vitro, and vice versa for the NPH lab. The image of electric wire led the NPHt to understand that a structural relation could be traced between carbon nanotubes and neurons: both materials conduct electricity.

> 'At least at the beginning, I think that it was easier for them (neurophysiologists) to understand, because of the assumption that nanotubes are like tiny electric wires connecting two neurons that communicate through electric signals.' (CHt)

The *electric wire* metaphor was triggered by the superficial similarities recognition and developed around the reconstruction of a deeper, structural one, which made salient a common feature of the biological and artificial material, the electrical conductivity, and suggested a common functional property of both: conducting the electrochemical signal of neurons. Hence the *electric wire* metaphor prompted the scientists to infer that nanotubes might act as substitutes of the dendritic connection.

The idea of exploiting the common relation – the connectivity property – of the two materials was pivotal in the first two years of the research, namely in the setup of the first experiment. The first experiment was created on the basis of neurophysiologic prior experimental results according to which neuronal tissues placed apart tended to grow more towards each other than in other directions. Therefore they tested the hypothesis that nanotubes like 'tiny electrical wires' might connect two slices of neuronal tissue set far enough not to be able to reconnect otherwise. The preparation work was divided between the two laboratories. The CHt functionalized nanotubes by making three adjustments to transform them into a biocompatible compound, adhesive to glass and non-water-soluble. These adaptations that yielded two within-discipline publications for the CH group. Then, the CHt provided this compound to the NPHt who was in charge to conduct in vitro experiments in her laboratory.

Such experiments failed. Following an established neurophysiologic experimental approach, a layer of protein material was put between neurons and carbon nanotubes, however it impeded the electrical signal to flow from neurons to nanotubes. Therefore, for a successful result, it should have been dropped. The *electric wire* metaphor, while successfully inspired the first joint experimental attempt, and the CH's work on nanotubes to finalize them for the experiment, they it did not help to fully reframe the standard experimental procedures used in NPH experiments.

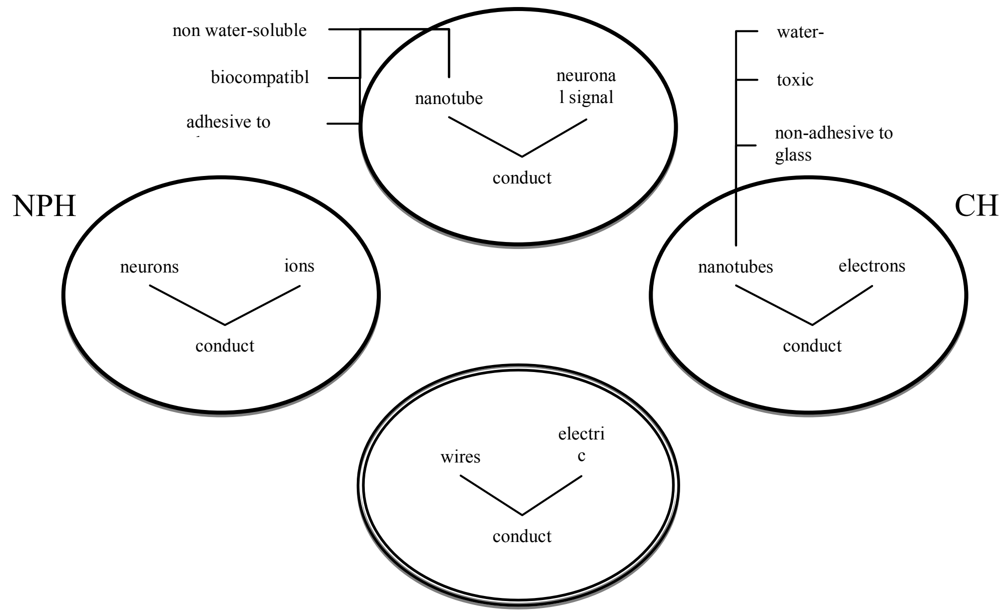
The 2005 experiment and the *scaffold* metaphor (metaphor 2)

In 2003 after the first experiment, the two researchers applied for a regional grant, committing themselves to a joint research project to study nanostructures of carbon and neural circuit formation. When the grant was awarded, they designed the new experiment drawing on previous results, specifically trying to understand what were the obstacles. This aim led the scientists to intensify their effort on the project and the frequency of their interaction with different solutions: joint-seminars and lectures, and the appointment of a CH graduate student to the project. Exchanging reading materials and increasing their interactions facilitated the identification of the material that insulated neurons from nanotubes.

Due to his longstanding research expertise, CHts was aware of the potential of the carbon structure as platform onto which developing a new generation of innovative medical therapies.

Nanotubes had a set of properties that allowed a new association within the CHt's mind. Nanotubes' have a porous and fractal-like configuration, constant along their length and apt to be chemically functionalized with molecules in order to change their properties. This feature combined with a proven compatibility with physiological conditions makes them potential devices to create nano-scale prosthesis, and adapt to be the 'scaffold' for neuronal growth and axonal regeneration. This *scaffold* metaphor was mentioned in an article of the CH group published in 1998 in which they refer to fullerene, the genitor of nanotubes, as potential scaffold for tissue growth, see Table 1. Thus due to the *scaffold* metaphor, some ideas from the broad knowledge domain of "tissue engineering", a hot topic in chemistry, especially with respect to carbon nanotubes (Harrison, & Atala, 2007), were brought in at the beginning of the second experiment.

Figure 4. Metaphors: the generic space and additional requirements



'[H]e (CHt) initially thought that nanotubes could be the platform which could direct the neuron's growth. Many studies in the neuronal regeneration share this idea of scaffold ... he had this idea because he worked with peptides and he knew he could steadily hook neurons to nanotubes which have an enormous surface. Since they are tubes, cylinders, they have a very interesting ratio of exposed surface which can be functionalised. He had this idea.' (NPHt)

The new experiment was developed on the basis of the *scaffold* metaphor by the chemistry laboratory, which brought in a different experimental approach, the removal of intermediate layers between neurons and nanotubes, and boosted the activity by involving a new young researcher.

The new experimental setting designed one year later consisted in neurons deposited directly onto a layer of carbon nanotubes in contrast with the prior experimental operative procedures that was one the main causes of the contradictory results: the two materials did not "speak to each other" (as they were interpreted by the NPHt). Electrical activity was then recorded. On a control glass, neurons were deposited directly on the borosilicate glass. Results showed a boosted neuronal activity in presence of nanotubes. These findings surprised both scientists and yielded the first joint publication in 2005.

"In the long term, our results will prompt the development of new tissue engineering strategies." (NPHt – Pr2Nph05).

The interpretation of results was framed under the lens of the *scaffold* metaphor of neuronal growth and regeneration to re-establish the connection after spinal injuries. Furthermore the new experimental setup required the advancement of the technique of explants of neurons (NPH) and the control of layer of nanotubes' thickness (CH). Both advancements were crystallized in the joint published work, as the experiment had successful, although puzzling, results.

The 2008 experiment and the *percolation* metaphor (metaphor 3)

The increased neuronal activity when neurons were coupled with nanotubes was puzzling because it was incompatible with the simple picture of dyadic relationships among neurons, implied by both metaphors of the *electric wire* and of the *scaffold*. Therefore the two scientists needed to look for new competences outside their own disciplinary domains. They decided to broaden the theoretical background of the team by drawing on a different discipline: electrical engineering. Thereby, in the 2005, the NPHt contacted the ENGr who got involved in the project with the role of developing mathematical models to predict the neuronal electrical activity.

> "At that time she (NPHt) had the curiosity to know what in the nanotubes enhanced the neuronal [electrical] activity and, naturally, we (chemists) wanted to understand what [was the] quality of the nanotubes we were using [that provoked that effect] in order to favour that [electrical] activity." (CHt)

The group thus worked on the understanding of the rationale of such puzzling phenomenon. To understand the cause of the boosted electrical activity, chemists' and neurophysiologists' skills, tools and knowledge bases were not sufficient. Since the phenomenon pointed to the electrical properties of nanomaterials, recruiting knowledge resources capable to model the interaction at the interface was crucial. The neurophysiologist recognized that integrating a new competence, bringing into the project a complementary theoretical perspective and methodological approach, was necessary to understand what happens at the nanoscale. Also, the CHt felt the need to move from a "phenomenological" (in the words of CHt) leading metaphor to a better specified frame that could act as a full model of the process, generating quantitative predictions and not only qualitative inferences. Recruiting the ENGr trained in the simulation of neuronal circuits provided the key source of knowledge to accomplish such step ahead.

A new research project was funded by the European grant (2006-2009) and the team was formed with the aim to develop implants, which may repair damages at the central nervous system. It encompassed new knowledge bases among which engineering was central in the design of experiments.

A new language orchestrated the activities of members directed by mathematical formulas and theoretically-driven by the ideas of ENGr and his small group, emphasizing the role of electrical circuits and networks of connections. The trigger for a new round of experiments came from such background.

> ‘I found an article by Kirkpatrick, a very famous physicist, who wrote an article in the 70s on percolation and electric conduction... what percolation really means? He basically talked of a resistive lattice and equivalent circuits, resistors, he used the elements I was comfortable with. And he showed how there is an electric path between two distinct points’. (ENGr)

Microscopy images, see Figure 3, provided the material context where such an intuition of a discontinuous interaction between materials is visible, therefore they support a deeper structural similarity between the percolation theory idea and the combination of artificial and neuronal graft. The new visionary idea brought by the metaphor of *percolation* was the passage of ions between any two points of the organic and the artificial layer.

> ‘Probably only the word percolation allowed me to see the same electron microscopy images in a different way. Those nanotubes touch each other, thereby I can imagine there is an electric path between any two points in the network.’ (ENGr)

The ENGr imagined the interaction between the neuron and a batch of carbon nanotubes like an ‘electric wire sitting on other electric wires’. Moreover this electrical metaphor was enhanced by the new conception of nanotubes that described them as ‘a dispersion of tiny wires, infinitesimal needles, electrically conductive’ (ENGr – Au11Eng08) and such image was allowed by the view of new images.

Through the metaphor of *percolation*, the two compounds were treated as homogeneous layers that might leak electricity at any point, therefore the electric current did not necessarily stream in one determinate direction. To test the hypothesis of the shortcut of current that returns to where it started, after leaking through the carbon nanotube substrate, a sequence of intermediate steps was required. The development of a way to describe phenomena under the language of electrical equivalent led to a mathematical formulation of possible interactions and, therefore, to a model describing the coupling between the neuron and nanotubes. Such an intertwined interaction between theoretical modeling and experimental data triggered by the *percolation* metaphor, required a tighter interaction between the two ENG and NPH laboratories. This implied a joint planning of the experimental design and controls and was one of the main problems to overcome in this phase of the project.

> '[T]he fact that it worked surprised me, because I thought: this (hypothesis) is science-fiction, this is a cartoon. It is not possible, there must be other explanations. And recently, few weeks ago, the NPHt carried out another experiment and it seems that another hypothesis that made me sleepless will be rejected.' (ENGr)

The outcome of the electrical measurement performed on the whole neuronal cell had to be tuned with the results of a computer simulation generated by a theoretical model developed by the ENGr.

The metaphor of *percolation* produced important results that confirmed the hypothesis of a leaking current from the neuronal membrane via nanotubes back into the soma, where it originated. Such work yielded the group's most relevant publication and a first theoretical model of interaction between the neuron and nanotubes.

DISCUSSION

In this section, we discuss the research findings addressing the research questions on how metaphors emerge in a context with multiple knowledge domains, how they affect action and what is, in turn, the effect of actions on metaphors. With respect to the extant literature, our accounts provide new ground to contribute on the theories of metaphors' role in knowledge combination processes.

The emergence dynamics of metaphors in interdisciplinary context

Our research findings show how metaphors emerge. From a point of view of metaphor generation dynamics, it seems to occur in two stages. In the first stage, bridging of distant disciplinary domains is based on the recognition of surface similarities, triggered by images. In a subsequent stage, the emergence of a common structure, i.e., the generic space, helps to detect even structural similarities.

The Electric wire Metaphor. Through the selective analysis of different microscopy images produced at the same scale of magnitude in each lab, the two scientists were surprised by the resemblance of the two images, and guided by their prior scientific background (Styles, 1997) they recognized surface similarities between the two objects: analogous branching structure, roughness of the surface, and elongations with similar caliber. Such morphological properties of nanotubes and neurons trigger a between-domains mapping and the reconstruction of a generic space that draws on electricity domain. This leads to a more in-depth analysis and to the recognition of a higher level (structural) similarity around which the first experiment is generated: the electrical conductivity.

Similarity detection and cross-domain mapping is the first step of a metaphoric thought (Wolff and Gentner 2011). To make sense as a metaphor, projections of parts of the input domains of chemistry and neurophysiology are blended together through the common ground of the generic space of electricity.

At the same time the selective nature of the *electric wire* metaphor hides some relevant problems relative to the integration of organic (neurons) and inorganic (nanotubes) objects. Thus while on one hand it inspired the project stet up showing the two objects' similarities, on the other it was not sufficient to design a successful experimental approach.

The Scaffold Metaphor. After the disciplinary work of adaptation made during the first phase of the collaboration, nanotubes present new properties matching those of a scaffold for tissue regeneration. As shown in Table 1, prior knowledge of scaffolds, associated to carbon molecule genitor to nanotubes, is already part of the CH group's background. Similar morphological attributes between the scaffolds and carbon nanotubes are matrix-wise, biocompatible meshes with a porous surface. In the domain of tissue engineering, scaffolds are used to support tissue formation or regeneration, thus scaffold help re-establishing the function through the re-generation of the tissue. This idea is compatible yet different with the aims of the second experiment that aims to verify the functional properties of neurons over the carbon scaffold to eventually provide functional, instead of physical, reconnection between injured neuronal tissues. The common relational structure, generic space, is the artificial surface that provides support to nervous tissues.

The Percolation Metaphor. Microscopy images at different magnification show discontinuous and frequent junctures between the surface of a neuronal dendrite and that of a nanotube, which is rough enough to pierce the organic material in different points. The need of a deeper understanding of the puzzling boosted neuronal activity that derived from the research advancements drives the search of new resources and ideas. The new *percolation* metaphor exploits surface similarities that consider nanotubes and neurons as a homogeneous domain and associate them to the knowledge domain of electricity. On the one hand there is a network of junctures between neurons and nanotubes (what the ENGr mentions as 'electrical wires sitting over other electrical wires'), on the other hand the theoretical framework of percolation theory that speaks about the connections in resistive lattice. Such a cross-domain mapping allow to draw on the general idea of the percolation theory, which becomes the generic space: an electrical circuit connects any two points, if there are enough connections – junctures – in the lattice.

Like in the first metaphor, the generic space is based on the electrical domain, from which through an enriched cognition and a series of highly magnified images the group is able to extract a new generative metaphor: the *percolation* one that in some ways comprises the past one of the 'tiny electrical wire'.

Findings show how the emergence of metaphors stems from the recognition of surface similarity that favors the identification of a common relational structure across domains. The role of images is central in this process. They help to map distant concepts and trigger structural similarities. Findings also show that metaphors are organized around the generic space.

Centrality of images and generic space

Our findings shed light on two key aspect of metaphors emergence in a interdisciplinary contest. First, research results show how images provide the ground for similarity recognition, overcoming the difficulties of bridging distant disciplinary domains that do not share a common language. Secondly, findings suggest that in the emergence of metaphors, the identification of an abstract generic space is a key step allowing the definition of a common ground for concepts and for actors' communication and action.

Different disciplinary communities of practices, well represented by each scientific group, see Table 1, can overcome boundaries by means of shared representations, such as images, that help scientists to translate ideas, mediate interaction by providing a tangible context (Bechky 2003; Ewenstein & Whyte, 2009; Henderson 1991). Yet, images serve another important function: they facilitate metaphoric thinking. This phenomenon occurs through a complex series of mental steps. First, images make apparent the nonobvious (cf. Shane, 2000) links between domains, by providing the context to match a familiar knowledge background with visual aspects of an unfamiliar one. This occurs because images are effective means to retrieve knowledge stacked in individuals' memory (Keane et al. 1994), providing the ground for the vision of morphological characteristics of unfamiliar domains, that are interpreted by means of that knowledge. Recognized surface similarities drive the establishment of further mental associations between objects: a symmetric mental process of similarity detection at the initial stage of the metaphor generation (Wolff and Gentner 2011). Secondly, in a later stage, when surface similarities are established, common structures between inputs emerge. The identification of the common structure is central in the process of metaphor generation, and it becomes a cognitive resource that might be available as a reference source domain eventually for solving future puzzles and give structure to future metaphors and solutions. As the cognitive science literature stresses, structural similarities tend to be cognitively demanding in absence of multiple cues or a guidance (Catrambone, & Holyoak 1989;

Gentner 1989), therefore, in absence of evident surface similarities, the emergence of a generic space is hindered.

In summary, images are the tools that provides interpretable signs for scientists to make sense of unfamiliar domains, and specifically images facilitate a correspondence-based emergence of metaphors (Cornelissen, 2005) by allowing the recognition of similar attributes across domains. These similar attributes in turn facilitate the emergence of a common generic space providing the structure on which new conceptual combination can be grounded. This sustains the revelation of subsequent and more profound – and at times successful – transfers of inferences and relations between domains. Therefore images trigger a creative process that starts with the identification of bridges across distant domains, goes through the generation of a generic space and a relational structure and sometimes it is completed by the adoption of inferences from the source domain into the target.

Table 5 – The generic space and the cognition of input domains

The generic space and the cognition of input domains

	Metaphor 1 - tiny electrical wires		**Metaphor 2 - scaffold**		**Metaphor 3 - percolation**	
Generic space	small caliber wires that bring the electricity from a source to a recipient which is not connected otherwise		artificial matrix-wise structure that is porous and allows the attachment of biologic material for tissue regeneration		a high probability of connections between nodes in a resistive lattice allows the percolation of electricity between any two points in the lattice	
	properties		properties		properties	
	before the experiment	after the experiment	before the experiment	after the experiment	before the experiment	after the experiment
Nanotubes	conductive and semiconductive small caliber wires; toxic; non adhesive to glass; water soluble.	conductive and semiconductive small caliber wires; biocompatible; adhesive to glass; non-water soluble.	properties mentioned after the experiment of Metaphor 1; random thickness of the deposit.	conductive and semiconductive small caliber wires; biocompatible; adhesive to glass; non-water soluble; controlled and thin thickness of the deposit	all of the properties mentioned before and in contact with neurons, they display contact ad distant and disconnected points.	same as before
Neurons	small caliber conductive elongation with a matrix protein as a support.	same as before	same as before	small caliber conductive elongation with no matrix protein support.	same as before	same as before

The impact of metaphors on actions

The strength of metaphors to provoke and drive collective actions has been stressed for the capacity of giving familiar and legitimate frames to new contexts (Hill & Levenhagen 1995; Cornelissen, & Clarke 2010; Cornelissen et al. 2011) that drives action towards new directions. We provide further evidence and extend the prior literature by showing how metaphors impact decisions and actors' networks.

As an organizer of the integration and combination of knowledge domains, the metaphor creates a new multi-disciplinary shared interest upon the creation of an experimental artifact, a common field in which practices and meanings do not necessarily merge but can still be coordinated. While experimental practices in neurophysiology and chemistry remain to some extent opaque to each other during the project, metaphors organize a common language allowing to conceive and evaluate experiments performed separately in each lab.

Metaphoric conceptions drive the disciplinary work of scientist by affecting the experimental setup. The *electric wire* metaphor triggers adjustments in the physical properties of materials and in experimental practices to leverage the common conductive functionality of neuronal tissues and nanotubes. At the earliest phase of the collaboration, nanotubes were toxic for neurons due to traces of heavy metals. Moreover, they had to be steadily hooked to the experimental glass not to harm neurons. Ultimately, nanotubes were also water-soluble and they disappeared when combined with the organic solution containing neurons. Disciplinary and experimental adjustments were necessary: nanotubes were depurated and made non-toxic, and adhesive to glass.

In the second phase, the *scaffold* metaphor suggested the opportunity to merge the two materials instead of having a layer in between, which guided the redesign of the new experiment and further disciplinary work by both the NPH and the CH laboratories. In order to allow the superimposition of the two materials, and the identification of neurons in the mesh, the CHt controlled the thickness of the nanotubes' layer which had to be apt to allowing visual count of neurons and the ability for the experimental hand to reach and measure the neuronal activity by means of the proper electrical recording pipettes. Meanwhile the NPHt modified the explantation technique removing the matrix protein material which prevented the functional communication between neurons and nanotubes.

However, even the method of the experiment was not immune to the effects of the metaphoric conception. To exploit the connecting idea of the *electric wire*, two layers of neurons were separated to see whether nanotubes may transfer the electric signal between them. In the third experiment, exploiting the concept of *percolation* to infer electrical shortcuts in the network, the experimental results have to interact with computer models since electrical measurements performed by the NPHt must be tested against the results generated by a computer simulation developed by the ENGr.

Metaphors have an impact not only on experimental adjustments, but also on the network of resources to mobilize and recruit, *in primis* researchers. For instance after the first successful experiment, the CH and NPH knowledge background is not sufficient to explain enhanced neuronal activity, therefore the two scientists look within their immediate social network for actors who may help framing the puzzle under a different perspective.

The impact of actions on metaphors

Findings show how-metaphors not only impact actions, but in turn are affected by research puzzles and advancements; metaphors change their role and their functions, give heuristic contributions and are also rhetorically used to disseminate results to multiple types of audience. Actions have a two-fold impact on metaphors. On the one hand, scientists'actions modify the cognitive and material resources of the team, enriching the knowledge endowment on which the original similarity recognition could be anchored and metaphors could emerge. On the other hand, during the project scientists may discover that a metaphor is loses its heuristic power. A metaphor can be discarded in terms of heuristic guidance in the process of scientific discovery, but it can still survive with a different function and for different targets, as it becomes a powerful rhetoric means in external communication, with sponsors, academia, and through the traditional press media. Furthermore, the metaphor is still a cognitive asset that the group can use in successive phases (for example the percolation metaphor builds on the idea of the network of overlapping electric wires).

To show the effect of actions on the cognition and therefore on subsequent metaphors, we consider the final part of the blend, which is the "elaboration" process (Fauconnier and Turner 1998), in which a mental (and material) simulation is run according to the metaphoric (experimental) rules that may generate results, puzzles and further thinking. Experimental actions modify material properties and scientists' cognition of the input domain, and since metaphors emerge on the basis of the latter, actions have a direct consequence on metaphoric production. For example, in the second stage, the transformation that carbon nanotubes had previously undergone making them biocompatible, non-water-soluble, and adhesive to glass enhanced their similarity with scaffolds, thus triggering the emergence of the metaphor.

Also changes of knowledge bases modify the endowment of theories held by the members, thus enabling members to draw on a broader (and deeper) set of domains and causal relations. One clear example is the third metaphor that stems from a deeper knowledge of electrical engineering. Scientists nest the first metaphor of the nanotube as an electric wire in a larger context of metaphors to produce a more complex mapping across multiple domains. In this phase, a neuron is also a tiny electric wire which sits on other electric wires – nanotubes – thus creating an electric network. And the electric network calls for a mapping with the domain of percolation, namely with the resistive lattice, used in Kirkpatrick (1973). Thus the recruitment of a new scientist that broadened and deepened the knowledge available gave rise to set of metaphors that construct a new generic space.

Table 6 – The pragmatic effect of metaphors and the cognitive effect of actions.

	Effects of metaphor 1	**Metaphor 2**		**Metaphor 3**	
	metaphor on actions	actions on metaphor	metaphor on actions	actions on metaphor	metaphor on actions
Actions	adaptation of nanotubes	partial failure and discard of the *electric wire* metaphor; adaptation of nanotubes makes them fit the role of scaffold	appointment of a junior researcher to perform all phases of the experiment; adaptation of neurons to the experimental design	puzzling results lead to the involvement of a new knowledge base (ENGr); re-use of the *electrical wire* metaphor as part of the new metaphor.	adoption of a new measurement technique.

The further use of metaphors in the communication of results with explanatory (Dunbar 1997, 1999) or legitimating intents (Gentner & Markman 1997) has already been analyzed by the literature. Along this stream, the analysis of radio and press interviews of scientists confirms that metaphors recur abundantly to express processes that are complex and unfamiliar to the audience in an easy and evocative way. However, they are also part of technical and peer-attended discourses. As part of their conversation, metaphors are *lively* (Ricoeur, 2003) not just in the rhetoric of the group, but as a cognitive tool to explore the frontier of knowledge and to combine different cognitive domains. Under this lens and because of the actions that scientists make, *worn-out* metaphors (Ricoeur, 2003), that should add no novel meanings, become useful and provide new meaning if the cognition of domains is enriched and new inferences may be made.

Thus, the heuristic power of a metaphor may be reinforced and enhanced if members' endowment of theories is enriched. If *lively* is the metaphor in which new meaning can be attributed, in which the conventional "cultural usage" do not decide "on the figurative sense of certain expressions" (2003: 225), we must point out that it is the cognitive richness of inputs domains that determines when a metaphor is still lively.

Tying the heuristic power of metaphor to the cognition of domains, namely showing the cyclic link between cognition, metaphor generation with actions that, in turn, modify the cognition may fade the dispute (Cornelissen 2005; 2006; Oswick et al. 2002; Oswick & Jones 2006) on the most apt trope for knowledge production, as also the same metaphor can be regenerated and inform new theoretical extraction. Consistent with both the comparison model of metaphor (Oswick et al. 2002) and the interaction model (Black 1962; Cornelissen 2006), our finding show that recognized similarities are a first step of metaphoric thinking. They inform and (constrain) the extraction of a deeper relational structure between domains. This generates the metaphor that enables the import of inferences to test in novel environments, such as the percolation hypothesis tested in the novel setting of neurons and nanotubes. We show that the heuristic value resides in two phases. The first occurs with the work of adaptation made on the input space to allow the combination of inputs. And this allows disciplinary knowledge development. The second consists in the combination of input domains, in which occurs the production of cross-disciplinary knowledge.

Being dynamic the cognition of the input domains, different cognitive endowments may inform different or nested metaphors, thus extending their creative power. For this reason, we believe that more critical point is the ability to see structural similarities, as they are the motor of the action.

CONCLUSIONS AND LIMITATIONS

Our paper shows that metaphors can effectively be mechanisms for knowledge production and exploration at the frontier of different disciplines, and describe the process of emergence of metaphors, their heuristic value for both disciplinary and cross-disciplinary communities, their guidance for scientific actions in multiple disciplinary setting does not come without flaws.

Our work enriches the soil of the scholars who indentified metaphors and analogies as players for innovation breakthroughs and scientific discoveries (Dunbar, 1997; 1999; Gassmann, & Zeschky 2008; Knoor, 1980), although our focus has been on how metaphors contribute to the combination of diverse sources of knowledge bases and their transformation (Carlile, 2002) guiding the action of actors belonging to different scientific communities.

We extend the understanding of images as fertile objects for creating a common ground between different social worlds (Ewenstein & Whyte 2009; Henderson, 1991). Images are the ground in which similarities are recognized, thus, setting the stage for the generation of metaphors.

Some critics may say that we omitted to analyze in depth the structure of the metaphor. We decided to undertake a different course due to our research aim on metaphors' dynamics. However our evidence might provide some interesting insight for further research delving on the theoretical debate that separate the vision of those who believe that metaphors are not the most creative tropes for knowledge production, being constrained by similarities (Oswick et al. 2002; 2006), and others who claim that through metaphors it is possible to go beyond the sum of the domains (Cornelissen 2004; 2005; 2006).This argument is definitely worthy to dedicate new energies and efforts and we believe that our data may give interesting information.

Our research has some limits and two natural lines of evolution. Our qualitative investigation could be enriched and complemented by a quantitative analysis of the textual material produced along the project. Formal text analysis could help to find regularities in language used by researchers and to understand, with a more fine-grained lens, how metaphors evolved.

A second line of research should investigate how metaphors are embedded in a context of social relationships and artifacts. Leading metaphors, scientific tools and the social network of scientists mobilized clearly coevolve along the project. Their interaction deserves closer analysis through an ethnographic approach.

REFERENCES

Argote, L., & Ingram, P. 2000. Knowledge transfer: A basis for the competitive advantage in firms. ***Organizational Behavior and Human Decision Processes***, 82: 150-169.

Baldwin, C. Y., & Clark, K. B. 2000. ***Design Rules, Volume 1. The Power of Modularity***. Cambridge, MA: The MIT Press.

Bechky, B. A. 2003. Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Production Floor. ***Organization Science***, 14: 312-330.

Black, M. 1962. ***Models and Metaphors.*** Ithaca: Cornell University Press.

Black, M. 1977. More about metaphor. ***Dialectica***, 3-4: 432-457.

Carlile, P. R. 2002. A Pragmatic View of Knowledge and Boundaries: Boundary Objects in New Product Development. ***Organization Science***, 13: 442-455.

Carlile, P. R. 2004. Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries. ***Organization Science***, 15: 555-568.

Catrambone, R., & Holyoak, K. J. 1989. Overcoming con textual limitations on problem solving transfer. ***Journal of Experimental Psychology: Learning, Memory, and Cognition***, 15: 1147–1156.

Clark, H. 1996. ***Using Language***. Cambridge University Press, Cambridge, U.K.

Cohen, M. D., & Levinthal, D. A. 1990. Absorptive capacity: a new perspective on learning and innovation. ***Administrative Science Quarterly***, 35: 128-152.

Corley, K.G., & Gioia, D.A. 2004. Identity Ambiguity and change in the wake of a corporate spin-off. ***Administrative Science Quarterly***, 49: 173-208.

Cornelissen, J. P. 2004. What Are We Playing At? Theatre, Organization, and the Use of Metaphor. ***Organization Studies***, 25: 705-726.

Cornelissen, J. P. 2005. Beyond Compare: Metaphor in Organization Theory. ***Academy of Management Review***, 30: 751-754.

Cornelissen, J.P. 2006. Metaphor in Organization Theory: Progress and the Past. ***Academy of Management Review***, 31:485-488.

Cornelissen, J. P., & Clarke, J. S. 2010. Imagining and Rationalizing Opportunities: Inductive Reasoning and the Creation and Justification of New Ventures. ***Academy of Management Review***, 35: 539-557.

Cornelissen, J. P., Holt, R., & Zundel, M. 2011. The Role of Analogy and Metaphor in Framing and Legitimization of Strategic Change. ***Organization Studies***, 32, 1701-1716.

Dunbar, K. 1997. How scientists think: Online creativity and conceptual change in science. In T.B. Ward, S.M. Smith, & S.Vaid (Eds.) ***Conceptual structures and processes: Emergence, discovery and Change*** (p.461-493). Washington DC: APA Press.

Dunbar, K. 1999. The Scientist InVivo: How scientists think and reason in the laboratory. In Magnani, L., Nersessian, N., & Thagard, P. ***Model-based reasoning in scientific discovery***:89-98. New York: Plenum Press.

Eisenhardt, K.M. 1989. Theories from Case Study Research. ***The Academy of Management Review***, 14: 532-550.

Ewenstein, B., & Whyte, J. 2009. Knowledge Practices in Design: The Role of Visual Representations as 'Epistemic Objects'. ***Organization Studies***, 30 (1), 7-30.

Fauconnier, G., & Turner M. 1998. Conceptual Integration Networks. ***Cognitive Science***, 22: 133-187.

Fauconnier, G., & Turner, M. 2002. ***The Way We Think***. New York: Basic Books.

Fauconnier, G., & Turner, M. 2008. Rethinking Metaphor. In R. W. Gibbs, Jr., (Ed.), ***The Cambridge handbook of metaphor and thought***: 53-66. New York: Cambridge University Press.

Gärdenfors, P. 2000. ***Conceptual Spaces: the Geometry of Thought***. Cambridge, MA: The MIT Press.

Gassmann, O., & Zeschky, M. 2008. Opening up the Solution Space: The Role of Analogical Thinking for Breakthrough Product Innovation. ***Creativity and Innovation Management***, 17: 97-106.

Gentner, D. 1983. Structure-Mapping: A Theoretical Framework for Analogy. ***Cognitive Science***, 7: 155-170.

Gentner, D. 1989. The mechanisms of analogical learning. In S. Vosniadou & A. Ortony (Eds.), ***Similarity and analogical reasoning***: 199–241. Cambridge, UK: Cambridge University Press.

Gentner, D., & A. B. Markman 1997, Structure Mapping in Analogy and Similarity, ***American Psychologist***, 52 (1), 45-56.

Gibbs, R. W. Jr. (Ed.). 2008. ***The Cambridge handbook of metaphor and thought***. New York: Cambridge University Press.

Gilovich, T. 1981. Seeing the past in the present: The effect of associations to familiar events on judgment and decisions. ***Journal of Personality and Social Psychology***, 40: 797-808.

Goguen, J. 1999. An Introduction to Algebraic Semiotics, with Applications to User Interface Design, in Computation for Metaphor, Analogy and Agents. In Christopher Nehaniv (ed.), ***Springer Lecture Notes in Artificial Intelligence***, volume 1562 (p. 242-291), Springer.

Grady, J.E., Oakley, T., & Coulson, S. 1999. Blending and metaphor. In G. Stern & R. Gibbs (Eds.), ***Metaphor in cognitive linguistics*** (p. 101-124). Amsterdam: John Benjamins.

Gregoire, D. A., & Sheperd, D. A. 2012. Technology-Market Combinations and the Identification of Entrepreneurial Opportunities: An Investigation of the Opportunity-Individual Nexus. ***Academy of Management Journal***, 55: 753-785.

Hall, B. 2000. Innovation and Market Value. In R. Barrell, G. Mason & M. O'Mahony (Eds.), ***Productivity, Innovation and Economic Performance***:177-198. Cambridge: Cambridge University Press.

Hargadon, A. B. 2002. Brokering Knowledge: Linking Learning and Innovation. In B. M.Staw & R. M. Kramer (Eds.), ***Research in Organizational Behavior***, 24: 41-85, Greenwich, Conn: JAI Press.

Hargadon, A. B., & Bechky, B. A. 2006. When Collections of Creatives Become Creative Collectives: A Field Study of Problem Solving at Work. ***Organization Science***, 17: 484-500.

Hargadon, A. B., & Sutton, R. I. 1997. Technology Brokering and Innovation in a Product Development Firm. ***Administrative Science Quarterly***, 42: 716-749.

Harrison, B. S., & Atala, A. 2007. Carbon nanotube applications for tissue engineering. ***Biomaterials***, 28: 344-353.

Henderson, K. 1991. Flexible Sketches and Inflexible Data Bases: Visual Communication, Conscription Devices, and Boundary Objects in Design Engineering. ***Science, Technology, & Human Values***, 16 (4), 448-473.

Hill, R. C., & Levenhagen, M. 1995. Metaphors and Mental Models: Sensemaking and Sensegiving in Innovative and Entrepreneurial Activities. ***Journal of Management***, 21: 1057-1074.

Hsiao, R. L., Tsai, D. H., & Lee, C. F. 2011. Collaborative Knowing: The adaptive Nature of Cross-Boundary Spanning. ***Journal of Management Studies***, articles first published online.

Huber, G.P. and Power, D.J. 1985. Retrospective Reports of Strategic-Level Managers: Guidelines for Increasing Their Accuracy, ***Strategic Management Journal***, 6: 171-180.

Jensen, R. J., & Szulanski, G. 2007. Template Use and the Effectiveness of Knowledge Transfer. ***Management Science***, 53: 1716-1730.

Keane, M. T., Ledgeway, T., & Duff, S. 1994. Constraints on analogical mapping: A comparison of three models. ***Cognitive Science***, 18: 387-438.

Kirkpatrick, S. 1973. Percolation and Conduction. ***Review of Modern Physics***, 45: 574-588.

Knoor, K. D. 1980. The Scientist as an Analogical Reasoner: a Critique of the Metaphor Theory of Innovation. ***Communication & Cognition***, 13: 183-208.

Kogut, B., & Zander, U. 1992. Knowledge of the firm, combinative capabilities and the replication of technology. ***Organization Studies***, 3: 383-397.

Lakoff, G. 1993. The contemporary theory of metaphor. In A. Ortony (Ed.) ***Metaphor and Thought*** (p. 202-251). Cambridge, UK: Cambridge University Press.

Lakoff, G., & Johnson, M. 1979. ***Metaphors We Live By***. Chicago: University of Chicago Press.

Leonard-Barton, D. 1990. A dual methodology for case studies: synergistic use of a longitudinal single site with deplicated multiple sites. ***Organization Science***, 1: 1-19.

Leonardi, P. M. 2011. Innovation Blindness: Culture, Frames, and Cross-Boundary Problem Construction in the Development of New Technology Concepts. ***Organization Science***, 22: 347-369.

Levina, N., & Vaast, E. 2005. The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems. ***MIS Quarterly***, 29: 335-363.

Miles, M., & Huberman, A. M. 1984. ***Qualitative Data Analysis***. Beverly Hills, CA: Sage.

Miller, C. C., Cardinal, L. B., & Glick, W. H. 1997. Retrospective Reports in Organizational Research: a Reexamination of Recent Evidence. ***Academy of Management Journal***, 40: 189-204.

Morgan, G. 1986. ***Images of Organizations***. Beverly Hills, CA: Sage.

Nonaka, I., & Takeuchi, H. 1995. ***The knowledge creating company: how Japanese companies create the dynamics of innovation***. New York: Oxford University Press.

Nooteboom, B., Van Haverbeke, W., Duysters, G., Gilsing, V., & van den Oord, A. 2007. Optimal cognitive distance and absorptive capacity. ***Research Policy***, 36: 1016-1034.

OECD 2010. ***Measuring Innovation: a new perspective***, OECD Publishing.

Ortony, A. 1993. ***Metaphor and Thought***. Cambridge, UK: Cambridge University Press.

Oswick, C., Keenoy, T., & Grant, D. 2002. Metaphor and Analogical Reasoning in Organization Theory: Beyond Orthodoxy. ***Academy of Management Review***, 27: 294-303.

Oswick, C., Jones, P. 2006. Beyond Correspondence? Metaphor in Organization Theory. ***Academy of Management Review***, 36: 318-337.

Oswick, C., Fleming, P., & Hanlon, G. 2011. From Borrowing to Blending: Rethinking the Processes of Organizational Theory Building. ***Academy of Management Review***, 31: 483-485.

Ricoeur, P. 2003. ***The Rule of Metaphor: The Creation of Meaning in Language***. London, UK: Routledge.

Rindova, V., M. Becerra, & Contardo, I. 2004. Enacting Competitve Wars: Competitive Activity, Language Games and Market Consequences. ***Academy of Management Review***, 29: 670-686.

Schoen, D. 1993. Generative metaphor: A perspective on problem-setting in social policy. In A. Ortony (Ed.) ***Metaphor and Thought*** (p. 137-163). Cambridge, UK: Cambridge University Press.

Siggelkow, N. 2007. Persuasion with Case Studies. ***Academy of Management Journal***, 50: 20-24.

Shane, S. 2000. Prior knowledge and the discovery of entrepreneurial opportunities. -***Organization Science***, 11: 448–469.

Shin, J. S., & Pierce, N. A. 2004. A synthetic DNA walker for molecular transport. ***Journal of the American Chemical Society***, 126: 10834–10835.

Smith, A. 1776. ***An Inquiry into the Nature and Causes of the Wealth of Nations***. London: W. Strahan and T. Cadell.

Stigliani, I., & Ravasi, D. 2012. Organizing Thoughts and Connecting Brains: Material Practices and the Transition from Individual to Group-Level Prospective Sensemaking, ***The Academy of Management Journal***, 55, 1232-1259.
Strauss, A., & Corbin, J. 1990. ***Basics of qualitative research: Grounded theory procedures and techniques***. Newbury Park, CA: Sage.
Styles, E. 1997. ***The psychology of attention***. Hove, England: Psychology Press.
Sutton, R. I., & Hargadon, A. B. 1996. Brainstorming Groups in Context: Effectiveness in a Product Design Firm. ***Administrative Science Quarterly***, 41: 685-718.
Van Maanen, J. 1979. The Fact of Fiction in Organizational Ethnography. ***Administrative Science Quarterly***, 24: 539-550.
Wolff, P., & Gentner, D. 2011. Structure-Mapping in Metaphor Comprehension. ***Cognitive Science***, 35: 1456-1488.
Yin, R. 1984. ***Case Study Research***. Beverly Hills, CA: Sage Publications.
Zahra, S. A., & George, G. 2002. Absorptive capacity: A review, reconceptualization, and extension. ***Academy of Management Review,*** 27: 185-203.

2. MOVING BOUNDARIES: A MULTIDISCIPLINARY COLLABORATION THROUGH TEXTUAL NETWORKS

Claudio Biscaro

Ca Foscari University of Venice

ABSTRACT

This study shows a method that stems from the combination of textual analysis with network theory and it has been used to visualize disciplinary boundaries in a cross-disciplinary collaboration of scientists. Through the analysis of scientific abstracts cited by an article, along with the abstract of the cited articles, semantics is extracted by collecting words and word relations. After parsing texts and eliminating stop words, a document text matrix is created to enumerate the frequency of words for each document. Word relations are set to be equal to 1 when two words appear to be in the same abstract, 0 otherwise. A term to term adjacency matrix is constructed to display networks of concepts across abstracts. Disciplines are detected by maximizing the modularity of the network. Depending on the network structure, it is possible to show points of contact between disciplinary fields either by eliminating the highest degree centrality elements, when brokering nodes are peripheral in both clusters, or by maintaining all nodes when connections at the core concepts of a discipline. We analyze different centrality measures to identify words across boundaries. Betweenness centrality and random-path betweenness scores better identify nodes as conceptual bridges through boundaries. This method combines the analysis of texts with the statistical tools of network theory enabling the passage from a theory-building approach of the qualitative study of boundaries to a theory-testing one.

Keywords: text analysis, network analysis, semantic analysis, boundary objects

INTRODUCTION

> "I find that the communities' knowledge-sharing difficulties are rooted in their work contexts, which differ on the basis of their language, the locus of their practice, and their conceptualization of the product" [Bechky, 2003: 314]

> "'Surface, it is true, is a substantive in grammar; but it is not the name of a particular existent, but of an attribute." [H. H. Price, *Perception* (1932: 106)]

Sometimes, words and concepts inhabit different worlds and signify differently according to semantic associations. Other times common concepts are bridges between worlds and ground (Clark, 1996) communication and eventually cooperation.

Thus, one of the criticalities of multidisciplinary research, which is indeed an important way to advance the scientific state of the art (Amabile 1988, OECD 2010), is to find bridges across disciplinary domains. When links are not clear and ambiguity is not understood, collaboration across different knowledge domains may be problematic (Faraj and Sproull, 2000, Leonardi, 2011). The coexistence of different representations, paradigms, knowledge bases creates issues for the knowledge to be bridged or simply combined (Bechky 2003; Carlile 2004). Failure may cause important distress, because tracing causality across boundaries takes time and is costly (Hsiao and others 2011).

Multidisciplinary knowledge production entails changes in disciplinary knowledge boundaries. The idea of combination of diverse skills and knowledge sets can be tracked back to the origins of the modern economic theorizing in discourses about enhancement of efficiency (Smith 1776) and innovation (Schumpeter 1936). More recent contributions highlight the dynamics, the function of boundaries, and the communication across them (e.g., Carlile, 2004; Abbott, 1995; Star 1989).

The concept of the boundary includes the ideas of cohesion within the boundary and separation from what is different. To define coherent systems from the outside, systems of concepts and representations, hereafter ontologies, are socially and dynamically tied. Demarcation is then an important idea underlying the concept of boundaries (cf. Gieryn, 1999; Lamont & Molnar 2002), not solely because through it a sense of identity and similarity among members is developed (Epstein 1992), but also because the creation of independent disciplines may enhance the distance across them. Interestingly, creative actions can be pursued by tying different ontologies together by the action of brokers (e.g., Hargadon and Sutton 1997) and also by means of specific objects that inhabit more groups without merging their knowledge base (e.g., Star, 1989; Star and Griesemer, 1989). The emergence of boundaries across fields is qualitatively constructed by looking at the core of ontologies, identifying social practices, language, conceptualizations and tools (Star and Griesemer 1989; Bechky 2003). By such an approach, the core differences, rather than the boundaries are identified. At the boundary, or better, at the interface between two or more groups, there are placed objects that may affect the communication flow. Yet, little is known on how it is possible to exploit relations to identify both ontologies and what constitutes the boundary if any, and this might be due to the difficulty of identifying a good level of analysis.

To look at the dynamics of disciplinary boundaries, an apt context is the relatively recent scientific research field of Nanotechnology, where there is a slightly larger amount of cross-collaboration of disciplines (Schummer, 2004). Known for the converging technology label attributed by a seminal research report (Roco and Bainbridge 2003, p. x), in which authors speculated that in the future unity of science must occur for the sake of society, relevant bibliometric evidences (as shown by Shapira et al. 2010) show that the level of cross-collaboration in nanoscience is mostly fictitious (Schummer 2004). Schummer shows that disciplines are moving separately at the nanoscale, rather than integrating. However, the same scholar sees a higher trend of cross-disciplinary collaboration in such fields rather than in others.

Notwithstanding the novelty of the research field, it is still possible to identify long-lasting multidisciplinary collaborations in Nanotechnology, and therefore, following their trails, there is the opportunity to study the integration of knowledge bases, analyze boundaries across fields and their dynamics, and the transformation and adaptation of knowledge fields. Thus, thanks to the stack of scientific material such research collaborations produce and work with, concepts like boundaries and objects at those boundaries that so far, in social sciences, have been treated purely qualitatively from scholars (Bechky 2003; Carlile 2004; Star and Griesemer 1989) can be analyzed using different lenses to explore a virgin territory that is the automatic quantification and visualization of such boundaries. For such reason, on the trails set by other scholars (Callon, Law, & Rip, 1986; Small 1972), we adopt a mixed approach of textual analysis – co-words, paraphrasing Callon and others – and network analysis, to describe changes within a 8-year-long multidisciplinary scientific collaboration. We find that disciplinary networks of concepts share analogous properties with the structure of communities, therefore they can be identified by maximizing the network modularity and we also test measures of network centrality to capture concepts at the boundary. We show changes in disciplinary boundaries, growth and modification of ontologies, as well as disciplinary marginalization, we suggest that this technique could aid and strengthen interpretation of qualitative research results. Furthermore, by combining network theory with its statistical methods, the heuristic value of this method goes beyond the visualization aspects, because this approach might be used to refine theories on boundaries and to move to a phase of theory testing.

In the following sections, we first introduce the idea of boundaries in science and what is known about cross-boundary collaboration. We then discuss the method of textual analysis combined with network analysis to understand how to treat textual data in order to draw relations between words and how to address the problem of what lays in between groups of connected words. With this method, abstracts of scientific papers and the ones of their citations are parsed and connections are established. Different centrality measures are described to theorize which of them could work best and in which case. Results show that concepts and relations between disciplines change over time, proving that what can be an almost obligatory intermediate concept at a point may lose much of its betweenness power in later stages of a collaboration.

BOUNDARIES, DEMARCATION AND CROSS-COLLABORATION IN SCIENCE

The notion of boundary is predominantly topogical. A boundary defines and divides, it creates an inside and, indeed, an outside. The attention on this topic within social sciences has spanned several fields of research such as social identity, class, ethnic, and racial issues, national and spatial identities, professions, as well as social research on science (see Lamont, and Molnar, 2002 for a review). Within a sociological perspective, scientific and professional boundaries (Parkin, 1974) are treated as a demarcation tool. They are needed to protect and distinguish scientists from charlatans and amateurs (Gieryn 1999). Credibility is maintained by a continuous work of expulsion, expansion and protection carried out through the boundary. Expulsion of others by claiming that they do not adopt proper scientific methods; expansion 'to monopolize jurisdictional control over a disputed ontological domain' (Lamont, & Molnar 2002: 179); protection against alternative outside powers, e.g., managers who try to step over and use the scientific authority for their sake. In this open context, groups compete and dynamically tie their boundaries (Abbott, 1995), thus defining (and distinguishing) themselves.

An interesting shift in the literature on scientific boundaries has moved attention towards knowledge production. Although knowledge production and communication across boundaries can be facilitated by the adoption of stable lexicons and standard methods within ontologies (Bowker, & Star 1999: 5), diverse ontologies – disciplines – may not have direct benefit from integration as their incentive systems, paradigm, lexicons differ, and it could be simply too costly for single individuals to mesh partly or entirely the two systems. In this case, it is possible to collaborate by means of shared concepts and objects that inhabit multiple ontologies (Star 1989; Star, and Griesemer, 1989). This idea has been explored in companies where objects often facilitate coordination and understanding between different communities (Levina & Vaast, 2005), often also transforming disciplinary knowledge (Bechky 2003; Carlile 2004).

In this trend of knowledge production, moving within disciplinary boundaries has been shown to be easier than trying to work across them. Bauer (1990) speculated that it is almost impossible for a scientitst to contribute to another disciplinary area, behavior that is also called boundary crossing, for the existence of different paradigms, for no or little incentives tie together tenure track development and cross-disciplinary collaborations (van Rijsoever, & Hessels 2011). One reason that hinders multidisciplinary research is represented by the disciplinary rules, praxes and expectations in peer review publishing processes that are better known by scientists of the same discipline (Perper 1989). However, different disciplinary paradigms do not impede the communication across boundaries as scientists borrow theories, methods, and ideas from other disciplines (Oswick, Fleming, & Hanlon, 2011 show an extensive borrowing within management theory), modifying the lexicon and enriching the field. Another form of knowledge production across boundaries is represented by boundary crossing behaviors of authors who publish in a journal of another discipline. In the case of co-authored articles, they represent cross-disciplinary collaboration, and by looking at affiliation networks of co-authors, links between several different disciplines can be traced (Newman, 2004). In the form of single-authored papers, instead, boundary crossing is increasing at least in social sciences (Pierce, 1999): the form Bauer (1990) believed to be the least plausible for the almost insurmountable obstacle of paradigmatic differences.

Collaborating across boundaries requires boundary spanning capabilities. Leaving apart the possibility of individuals to span across boundaries, while focusing on objects and concepts, the criticality of establishing some common ground has been recognized (Clark 1996), in order to interpret signals coming from different ontologies with a low cost of decodification and with high possibility of solving ambiguity. Objects (Bechky 2003) may also aid the creation of common grounds across ontologies to resolve among multiple interpretations. Objects and concepts work better as common ground for boundary spanning when they already inhabit both worlds, thus they do make sense to both disciplinary groups; otherwise, they may hinder or impede successful communication (Levina and Vaast, 2005).

As disciplines evolve, boundaries necessarily and constantly move, due to internal as well as external evolution and the direct external influence within the boundary caused by borrowing, collaboration, or boundary crossing behavior. However, on the one hand, scientific literature has studied phenomena occurring across boundaries by looking at networks of either co-citations or co-authorship, checking the mismatch of affiliation of author and discipline of the journal, thus giving insights on the structure of macrophenomena. On the other hand, it focused on purely qualitative research to understand how collaboration occurs at the micro-level. Literature often relies on solely qualitative data to study boundary phenomena. In the knowledge production literature, scholars studied objects placed at the boundary, neglecting the structure of boundary relations. This polarization leaves a theoretical gap in the middle both in terms of understanding of what is the dynamics of disciplinary boundaries and also methodologically, as there has been no relevant research on how to map boundaries thus far.

To cope with this lack of theory, we follow the historical path of a collaboration in Nanotechnology in which scientists are trained and belong to different disciplinary fields: neurophysiology and chemistry, and in a subsequent phase, also engineering. Scientists collaborate while keeping different visions. They work with different paradigms and tools. This distance is visible even to the untrained eye who may lack knowledge in chemistry, neurophysiology, and engineering. If this is the case, this method will help researchers to identifies cores and boundaries of different disciplinary fields, even when the knowledge of those is limited. In the next sections, we describe the foundations of the followed approach that combines text mining techniques with network analysis in order to understand the semantic relations between ontologies, tie them together to identify disciplines, boundary relations and their dynamics.

METHOD

Textual Analysis

Automatic textual analysis has recently emerged as a method of research over large corpora of data and it is the core of conferences of scholars and nowadays several publications adopt computational techniques to parse, retrieve and analyze texts written in natural language. Automatic textual analysis has been performed for disparate purposes: to extract keywords (Rose, Engel, Cramer & Cowley, 2010), to cluster and classify documents(Caldas, & Soibelman, 2003), for the analysis of political speeches (Hirst, Riabinin & Graham, 2010), and also for prevention of criminal events (Gianvecchio, Xie, Wu & Wang 2008) and many others.

In our case, we decide to investigates the relations between disciplines by looking at how concepts are distributed and to what other concepts they are connected across articles. Words and their relations are able to reveal similarity as well as change. For instance Murtagh and his colleagues (2009) track anomalies in the film script of Casablanca by constructing a document-term matrix that includes all words used in each scene in columns, and scenes in rows, and by performing hierarchical clustering to measure the distance among each scene. They are able to visualize the topical moments of the movie, the ones in which unexpected things occur.

Differences in semantics have been also considered on the basis of a common entity (Cenci, Pozzi and Borsacchi, 2010), such as in the case of the death of Michael Jackson, where fragments of text from Youtube's videos comments and newspaper articles have been collected and analyzed to detect differences between the acceptation in the press coverage and in the reaction of the video audience, showing significant differences in the meaning attributed by people and press.

Differences and classification are also used in the analysis of political speeches. By looking at term frequency normalized on the length of the speech, it is possible to unveil different frames underpinning different ideologies, not necessarily reflected in votes (Hirst, Riabinin, Graham, 2010). Therefore words express different systems of beliefs and ideology which may not be necessarily reflected in actions.

In scientific research, co-occurrences of words may reveal the structure of research within research specialities (Bahattacharya, Basu 1998; Callon, Courtial, & Laville 1991). Through co-occurrence analysis, relational representations of semantics underlying texts can be drawn and they are mathematically treated to construct visual representations using graphs: the basic component of networks. This technique has been adopted for exploratory analysis and give synthetic representations of concepts within different fields, such as network research, biological safety, medicine, biology , information studies, and other disciplinary fields (e.g., Callon et al. 1991, Looze, & Lemarie 1997, Ding, Chowdhury, & Foo, 2001). Moreover, the possibility of representing word relations in a network structure enables the possibility to study the dynamics of networks over time periods to study pattern, identify trends and identify new themes (Ding, Chowdhury, & Foo, 2001).

Disciplines, bridges and betweenness

The common approach of network theorists to study cross-disciplinary collaboration has been to look at patterns of co-authorship (e.g., Newman 2001, Newman 2004, Acedo, Barroso, Casanueva, & Galán 2006), neglecting the information contained in their output: scientific texts. In natural language it is possible to identify common bridges across disciplines, as most words are shared across communities. Most of these words are common words which do not belong to any jargon, therefore they must be excluded by the analysis to let cross-disciplinary conceptual bridges emerge along with the structure of disciplinary concepts.

Before looking for bridges, it is necessary to identify what is a discipline. A discipline could be thought as a more or less tight structure of concepts, methods and tools used. In a network of concepts in which concepts represents nodes, the edge is the co-occurrence of concepts in a chuck of text. If the weight of an edge is proportional to the co-occurrence between nodes, occasional co-occurrences will have weak ties, whereas when words are regularly in same texts, their relations will be stronger. In such a weighted network, modularity *Q* (Newman 2004: 056131-6) index can be used to measure the density of edges within a set of nodes with respect to those between other sets. The index varies between -1 and 1 and “values around 0.3 or more usually indicate good divisions.”

$$Q = \frac{1}{2m}\sum_{ij}\left(A_{ij} - \frac{k_i k_j}{2m}\right)\delta(c_i, c_j),$$

where $m = \frac{1}{2}\sum_{ij} A_{ij}$ is the number of edges in the graph, A_{ij} is the weight of the edge between *j* and *i*, k_i is the degree of *i*, and the function $\delta(c_i, c_j)$ takes values 0 when *i* and *j* belong to different partitions, and 1 otherwise.

In cross-disciplinary collaborations shared concepts that inhabit more worlds may lay at the frontier[5] of clusters, thus linking two or more ontologies. The removal of those concepts separate disciplines. Betweenness centrality (Freeman 1977, 1979) – or shortest-path betweenness (SPB), as we will use other measures that look at concepts in between – of a concept (node) is constructed to keep track of the number of geodesic (i.e. minimal length) paths in the network that go through a given node. Defined $g_i^{(st)}$ the number of geodesic paths between two nodes *s* and *t*, SPB through the

[5] We think of the frontier as the boundary that defines the inside from an outside. Not in terms of a network representation where position is not necessarily representative of a disciplinary boundary. In this network representation, highly interconnected nodes, which tend to appear as inside a larger group of nodes, could be unique links with other disciplines. They still represent our boundary, and they are objects-concepts that inhabit two worlds.

node *i* is the sum of all shortest paths that connect s and *t* and pass through the vertex *i* averaged over the total number of geodesic path between the two vertices *s* and *t* – defined as n_{st} – normalized over half of the possible paths.

$$b_i = \frac{\sum_{s<t} \frac{g_i^{(st)}/n_{st}}{g_{ij}}}{\left(\frac{1}{2}\right)n(n-1)}.$$

This index shows whether a word is necessary or could be easily substituted. The removal of nodes with highest score will most disturb the relation between other vertices (Newman, 2010), as they lie on the largest number of shortest paths, and could make the network fall apart into sets of cohesive – and coherent when applied to related words – components. However, being correlated with degree in most networks (Goh, Oh, Kahng, & Kim, 2003), the SPB score is positively correlated with the size of partitions the vertex bridges, thus yielding low scores when a discipline is represented by a few vertices.

In citation networks, high scores of SPB are interpreted as signal of interdisciplinarity – bridging across disciplines – of journals. Within the field of nanotechnology, for instance, Journal of Nanoscience and Nanotechnology and Nano Letters have relatively high values of betweenness centrality as they stand between journals of physics and chemistry (Leydesdorff, 2007).

However, few problems are coupled with the SPB measure, as it computes the centrality only through the geodesic path, giving low scores to nodes that fall off the shortest path between clusters. This measures entails that information have already a knowledge of the network and reaches its destination only through the shortest links. As noted by Borgatti when studying directed networks(2005), network structure affects the adoption of the measure of centrality. The adoption of the centrality measure should be tuned on the type of replication (or transfer) of information and the preexisting knowledge of the network by means of nodes. Knowledge of the entire network of concepts cannot be assumed in a interdisciplinary collective work, thus we need to explore also other measures for such undirected textual network.

Alternative measures of betweenness centrality

To our knowledge there is no literature on boundaries in using textual network , therefore there is the need to compare more measures of centrality to understand which best detects nodes at the boundary.

To overcome the assumption of betweenness centrality that implies that information knows where to flow and takes always (one of) the shortest route(s), alternative measures have been developed to highlight points which stand in between groups or are more important because they have control over other nodes. We look at three alternative measures of centrality: flow betweenness (FB) (Freeman, Borgatti, & White 1991), random walk betweenness (RWB) (Newman 2005) and power centrality (PC) (Bonacich, 1987).

With the metaphor of information exchange, node *i* is a mediator of information between *s* and *t* if it lays along a path connecting the latter nodes. To address the problem underlying in the assumption of the SPB and to understand where information flows, Freeman, Borgatti and White (1991) suggested a measure whose idea is that information saturates the capacity of each edge. Thus, nodes communicate through every edge at their maximum capacity. The maximum capacity is equal to the minimum cut capacity (Ford and Fulkerson 1962), following their min-cut, max-flow theorem. Therefore the flow-betweenness of a node *i* is the maximum flow through the node *i* over all possible *s* and *t*.

Be m_{st} the maximum flow from nodes *s* to *t*, and $m_{st}(i)$ the maximum flow from s to *t* passing through *i* averaged over all maximum flow between s and t

$$C_F(i) = \frac{\sum_{s<t}^{n} \sum_{t}^{n} m_{st}(i)}{\sum_{s<t}^{n} \sum_{t}^{n} m_{st}}.$$

We use the normalized measure of FB (Freeman *et al.*, 1991) which divides the score by the maximum possible value.

Not knowing whether the flow of information saturates the capacity of edges, while it may randomly move within the network until it reach its recipient or, better, until a discipline connects with the other, we compute also the RWB (Newman 2005: 42) which is appropriate in cases in which there is no knowledge about how information wanders within the network until it finds the destination. The RWB of a node *i* can be roughly defined as "the number of times that a random walk starting at *s* and ending at *t* passes through *i* along the way, averaged over all *s* and *t*." This is a probabilistic measure of betweenness as each unit of information has a probability of flowing through an edge proportional to the capacity of the edge with respect to all the capacity of the edges connected to *i*.

If information is transmitted to the linked nodes with a probability β at each step, EC (Bonacich, 1987) is derivable, as β reflects the degree to which far nodes are considered. When β = 0, power centrality of a node *i* is proportional to its degree. As β increases, the EC score of connected nodes are increasingly taken into account. Power centrality of a node *i* is then the expected number of times in which the information walks through a node *i*, averaged over all starting nodes. Small values of β emphasizes local structure, while large values take emphasize nodes that are linked to large set of nodes which are in turn linked to other sets of nodes, in a cascade fashion.

Data

The data used for this work are texts coming from the abstracts of scientific articles.

Following the experimental articles published by a group of scientists whose aim is to repair damages in the central nervous system by working with carbon nanotubes and neurons, we create collections of abstracts deriving from the one of their article and those of the cited ones. For example, if article A cites *n* articles, then a collection is made by $n+1$ abstracts.

We select three experimental articles for their pivotal role in the scientific collaboration under scrutiny and on this thread of developing technology. They are their first three experimental joint articles that are published in 2005, 2007 and 2009.

Table 1 - Summary of data

Main publication	2005		2007		2009	
Journal	Nano Letters		Journal of Neuroscience		Nature Nanotechnology	
# references with abstracts	11		23		30	
total abstracts	12		24		31	
	All text	After preprocessing	All text	After preprocessing	All text	After preprocessing
# lexemes	820	551	122 6	904	133 8	985
# tokens	248 3	1116	353 8	1831	497 8	2583

The analysis is focused on abstracts and specifically on words. Therefore books are excluded as without abstract. The first corpus is made by the abstract of the article published in 2005 and those of its 11 cited papers. In total, they account for 820 types of lexemes, which are different types of words counting separately all verbal forms and singulars-plurals. 2483 tokens, i.e., the total occurrences of words. After eliminating stopwords and preprocessing text (see §data analysis), lexemes are reduced to 551 and tokens to 1116. The second article considered, which is the one published in 2007, cites other 23 articles. The total number of different lexemes is 1226, which account for 3538 tokens. After the elimination of stopwords and preprocessing text, the total number of tokes decreases to 1831 in 904 lexemes. The third article, published in 2009, has 30 citation, but one short article has no abstract, therefore it is considered without content (and it will appear as number 5 in articles stemming from Cellot 2009, see §results). Overall 30 abstracts have 1338 different lexemes and 4978 tokens. After cleaning the text, lexemes become 985 and tokens 2583, see Table 1 for a short summary.

Data description: a quick journey in the nanotechnology collaboration

The first collection of articles stem from an experience in which two groups of scientists, neurophysiologists and chemists, who had never worked together, join to produce an experiment that will have unexpectedly promising results. In presence of carbon nanotubes, neuronal electrical activity is potentiated. Their first publication is dated 2005, on Nano Letters and marks the beginning of a new research collaboration and of a promising research field., they contact a third scientific group of bio-engineers, experts in modeling neural – and neuronal – networks.

In 2007, a third knowledge base of bio-engineering joins the formation with chemists and neurophysiologists and publishes a second experimental article in the Journal of Neuroscience, in which they make a first electrical model of the interaction of carbon nanotubes with neurons. Mathematical and electrical models provide a new lens to construct and test hypothesis and theoretically this means that they might understand what electrical properties of carbon nanotubes provokes the neuronal augmented activity.

This yields to an intuition which brings scientists to a new publication in Nature Nanotechnology in 2009. Supported (and triggered) by microscopy images, they see that the rough surface of nanotubes could pierce neurons discontinuously. They imagine that the current could leak from the neuronal elongations, called dendrites, into the carbon nanotubes, and maybe revert into the soma – the nucleus of the neuron –, where it originated. Thus modifying the firing rate of the neuron.

Data Analysis

Scientific abstracts are extracted through the ISI Web of Knowledge repository and, when not available, manually retrieved by the website of the journal. A software has been written to load and parse each word of a collection of abstracts and create a document-term matrix. Abstracts have been divided into three collections, see Table 1, from which three matrices stem. Such document-term matrices have words (lexemes) as columns and documents (abstracts) as rows with each entry representing the frequency with which a word appears in a document.

In the preprocessing phase, common words as articles, prepositions, and conjunctions are excluded before creating the matrix. To do so, we used the lists of stop words available online at code.google.com/p/stop-words/ which were integrated a specific list of words irrelevant for our purpose and rather common in abstracts, e.g., show, offer, prove, increase, depict, exhibit. Such lists are available under request. As the intent is to understand significant differences and similarities across disciplinary ontologies, the elimination of such words depurates articles from content that is not meaningful for the purpose of this work.

Documents are embedded in a metric space defined by the total number of words in a document-term matrix. Words are the attribute of the document vector, and their frequency represents the value of the attribute. It is possible to reconstruct the topology of information by looking at the hierarchical distance. Euclidean distance is an effective measure to study the dissimilarity across documents, and it is used because population of data is homogeneous (Murtagh, Ganz, & McKie, 2009), whereas the method of hierarchical clustering is well suited for displaying discontinuity across articles as it tracks differences. Clusters are created by iteratively joining the two most similar documents. Hierarchical clustering is performed with the Ward's method aiming to find the minimum distance to agglomerate documents. This passage is computed with full document-term matrices. Results are represented in Figure 1, where abstracts are enumerated starting from the citing article with label equal to 1. For instance, on the right part of the Figure 1 abstracts stemming by the article by Cellot and colleagues in Nature Nanotechnology 2009 can be grouped in three sets: from 6 to 4; from 7 to 10, and from 25 to 21. At the bottom of the figure the lines that stem from articles are joined if their Euclidean distance is low. Moving up in the figure, they join at a higher distance. It is also particularly visible that the two groups on the left are more similar to each other, while the group of abstracts on the right is distant from the others. Also Manhattan distance has been applied and it yields similar results as also shown in the study by Amine, Elberrichi, and Simonet (2010). Similar, but less readable and rich, results could also be obtained by computing cosine similarity between rows (abstracts) within the document-term matrices, we show cosine similarity just for the first corpus, in Table 2, as it does not provide further insights on data.

For the structure of scientific abstracts, linguistically essential, clustering may not be sensitive to different lexical choices. However, we controlled for it by cutting off the clustering all lexemes occurring 2 or less times, with the intent of controlling for the adoption of different linguistic forms. Hierarchical clustering showed similar results, with a lower overall distance across documents due to the reduction of significant amounts of tokens.

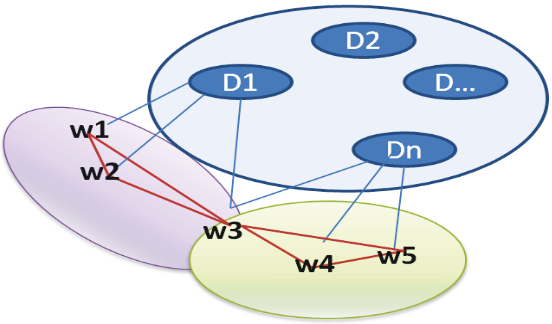
Figure 1 - The intuition behind the textual boundary. Documents are vectors and the words contained are the attributes with value one if is present, zero otherwise. This representation creates a document term (DT) matrix through which it is possible to extract a textual adjacency matrix (W). In the example, *table* is at the boundary between the food set represented in the words *red*, *apple*, and the surface set by *wood* and *surface*.

w_i = words, $i = 1, 2, \dots, m$

D_j = documents, $j = 1, 2, \dots ,$

DT =

Doc\W	Red	Apple	Table	Wood	Surface
Doc 1	1	1	1		
Doc 2					
…					
Doc n			1	1	1



$$w_{ik} = \sum_{j=0}^{n} w_{i,doc\,j}\, w_{k,doc\,j}$$

$$\mathbf{W} = \mathbf{DT}\,\mathbf{DT}^{T}$$

In order to create a network representation starting from a document-term matrix, we treat tokens as nodes – vertices – and we assume that link exist between any two words within an abstract, thus yielding undirected graphs. This assumption is sustained by the lack of negative sentences in our collections of documents (e.g., *a* is not *b*), and by the short size of each text. However, as the main interest is to understand the dynamics of the main concepts and their role within or across disciplines, words with less occurrences than a threshold are excluded from the analysis, thus leaving a reduced document-term matrix. We use different thresholds that vary according to the numerosity of abstracts and total number of words: 3 for the articles by Lovat and others 2005 on Nano Letters; 3 for Mazzatenta and others 2007 on Journal of Neuroscience; 7 for the article by

Cellot and colleagues 2009 in Nature Nanotechnology. At this level, we create an adjacency matrix containing relations among words from the reduced document-term matrix. The adjacency matrix contains all most used words of abstracts of each main article and its citations. For its construction, three steps are necessary. First, we eliminate information on frequency within each row-vector (abstracts) setting equal to one all positive occurrences, leaving equal to zero others. This operation on the one hand loses information about salience, but on the other it preserves relations between words and reduces the impact of the style of journal and writing which leads to different types of abstract, i.e., more or less detailed, shorter or longer. Second, the adjacency matrix is created by multiplying the reduced document-term matrix by its transposed. Third, self-edges are removed by setting the elements along the main diagonal equal to zero.

For each graph we compute over all nodes the average path length showing how many steps (how many edges-links) through the shortest path it takes from any node (word) to reach each other node. This gives idea of the cohesion of the network, but it is also sensitive to size. However, we control for size simply by means of the threshold that filters just the most used terms.

For visualization purposes and to eliminate *trivial* links between ontologies, as we are mainly interested on *hidden* relations, words with highest degree centrality, i.e., nodes with the highest number of edges, are removed from the network as they catalyze most of the words to them and tend to be too generic. The cutoff is calculated as 3/5 between the maximum and the minimum degree centrality scores to exclude the higher score vertices. In our case words like neuron and cell are always excluded. This cutoff point depends on the network structure and on the skewness of the distribution of degree centrality of nodes. If the network structure has relatively large clusters and nodes which broker across clusters that have relatively few edges, then it is applicable. Otherwise, in a structure where brokering nodes have many edges, they will be removed. In the last sets of abstracts (§Third phase), we retain all nodes, for the nature of the network that is tightly interconnected. For the network analysis and visualization, we use the both the software Gephi for one-mode networks and the package igraph for the bipartite networks (Csardi, Nepusz, 2006) running in the software R.

Graphs are then plotted, as in figure 7 – Cellot et al. 2009 Nature Nanotechnology. This figure exhibits a network of textual data. We maximize modularity with an algorithm that has been tested on other modular networks (Blondel, Guillaume, Lambiotte, Lefebvre 2008) and we partition nodes by color according to the community they belong to. Four measures of centrality such as betweenness centrality (Freeman, 1979) power centrality (Bonacich, 1987), flow betweenness (Freeman et al. 1991) and random-walk betweenness (Newman 2005) are computed to identify conceptual bridges. Betweenness centrality to give a size to nodes in the one-mode networks.

As a further step, to see beyond the visual representation of hierarchical clustering, we relate the abstract to their textual content creating a bipartite network, through an incidence matrix. The result is represented by bipartite textual networks and their origins in which scientific abstracts – numbers – are nodes connected to their related concepts.

RESULTS: DISCIPLINARY DYNAMICS

Telling the story of a multi-disciplinary collaboration can be aided by textual networks that show the dynamics of changing relations, the emerging of concepts and the different level of disciplinary integration over time.

The beginning of the collaboration, integration of two disciplines

The first publication of 2005 in which the world of chemistry and neurophysiology are combined in order to try to couple carbon nanotubes (chemistry) with neurons (neurophysiology) shows that the two worlds are still rather heterogeneous, distant, with few points of contact. The hierarchical clustering, in Figure 2 on the left, displays a group of similar articles (from 1, 2, 3, 4, and 6) that are more similar to each other than the others. They are chemically oriented, as shown in Figure 3, as they swirl around the chemically specific concept of carbon nanotubes – functionalization and others – and have some connections the article number 2, which belongs to a more medical-pharmacological world: the concepts of orthopedic and probes characterize it.

Figure 2 - Hierarchical clustering of the text in the sets of abstracts.

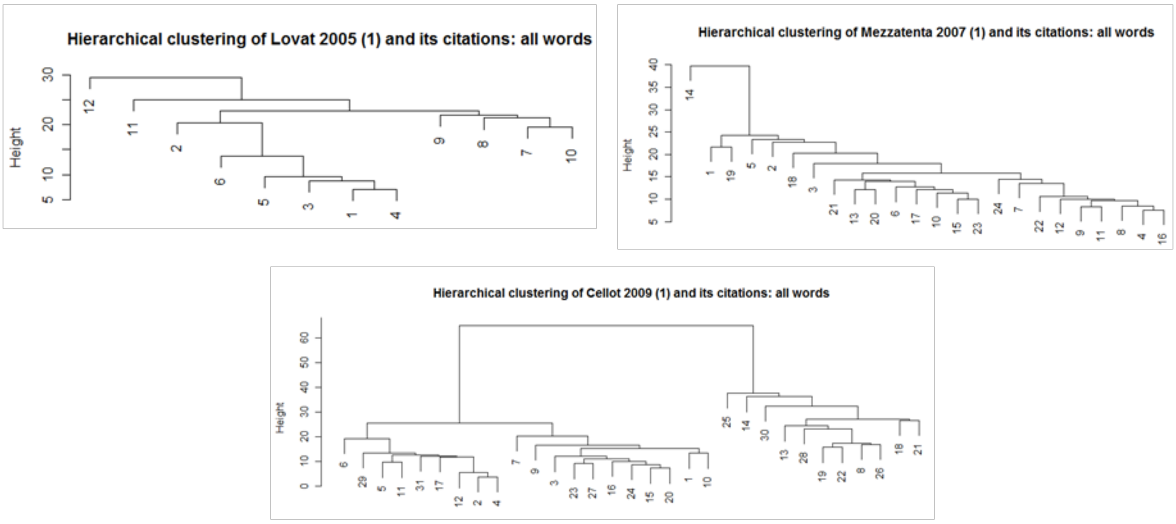
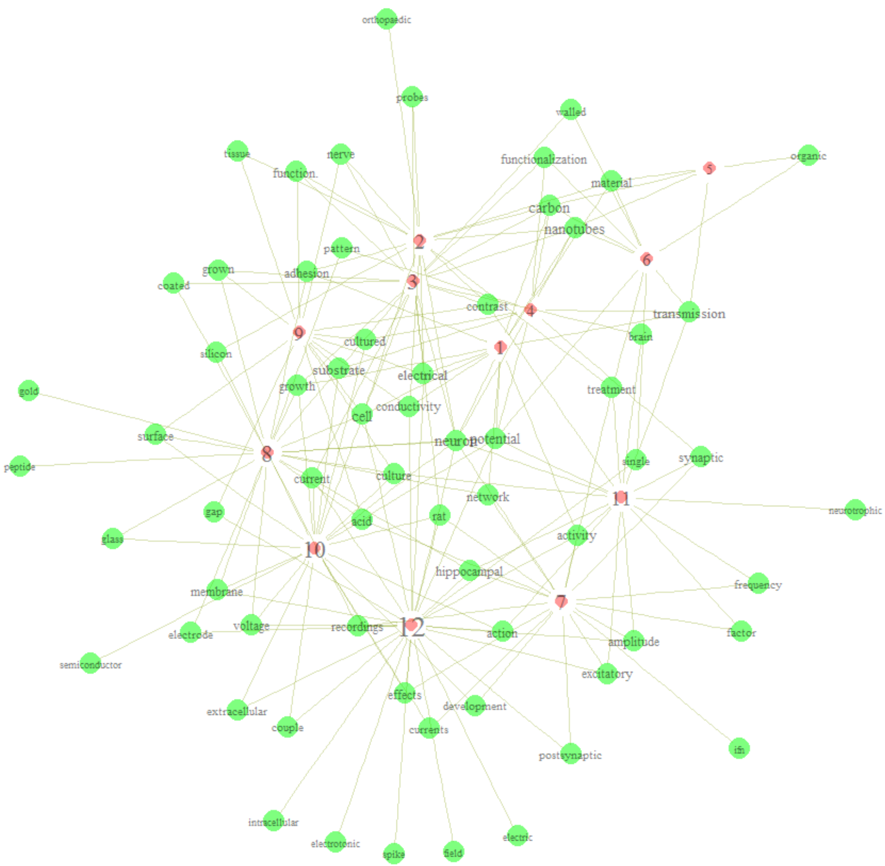


Figure 3- Bipartite network: abstracts and concepts stemming from *Lovat et al. 2005* and its citations



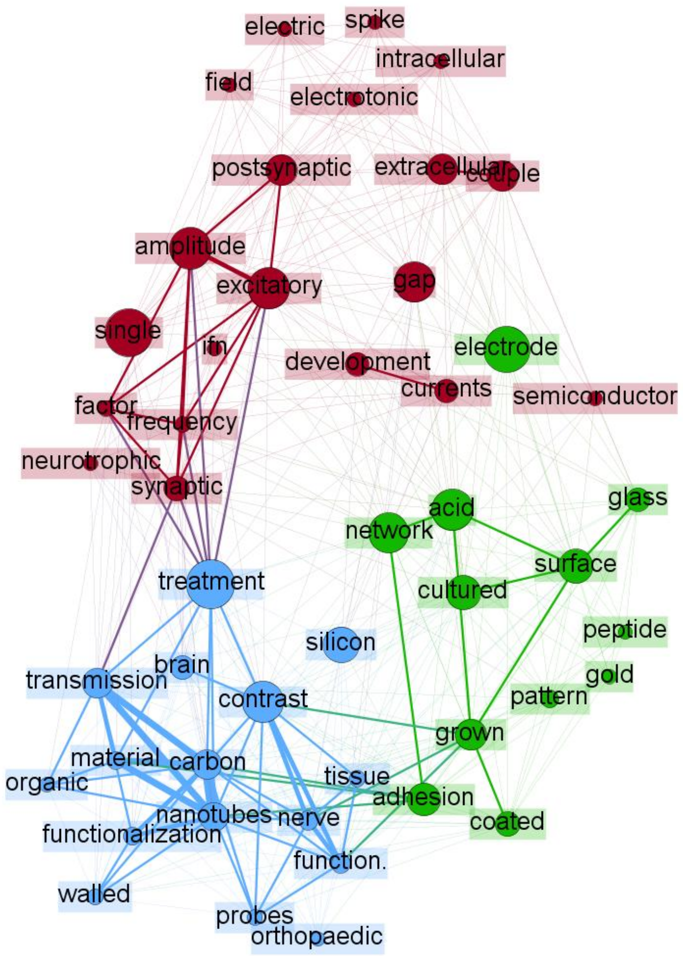
Cosine similarity calculated between documents of the document-term matrix indicates higher similarity within articles 1-6, as Table 2. However, cosine similarity does not give a qualitative information on the relations among articles. As cosine similarity is a symmetric index (i.e., the similarity between document 1 and 10 is equal to the similarity between 10 and 1), we provide a table that displays similarity measures computed on a reduced document-term matrix that takes into account lexemes repeated more than twice on the white background, and computed on the original document-text matrix on the grey background.

Table 2 - Cosine similarity in the text of abstracts from *Lovat 2005, Nano Letters* and those in its references. On the upper triangle, similarity is computed on a reduced document-term matrix

	1	2	3	4	5	6	7	8	9	10	11	12
1	--	0.68	0.77	0.76	0.62	0.56	0.26	0.25	0.12	0.30	0.13	0.21
2	0.54	--	0.63	0.57	0.34	0.33	0.24	0.24	0.28	0.31	0.08	0.18
3	0.61	0.50	--	0.75	0.63	0.63	0.19	0.22	0.19	0.29	0.11	0.12
4	0.54	0.40	0.55	--	0.57	0.58	0.24	0.27	0.04	0.33	0.15	0.15
5	0.44	0.23	0.45	0.36	--	0.74	0.00	0.02	0.00	0.00	0.02	0.00
6	0.46	0.27	0.52	0.44	0.55	--	0.01	0.00	0.00	0.01	0.03	0.04
7	0.19	0.17	0.14	0.15	0.00	0.01	--	0.16	0.03	0.29	0.31	0.17
8	0.20	0.19	0.17	0.20	0.01	0.00	0.12	--	0.18	0.30	0.05	0.16
9	0.09	0.20	0.13	0.02	0.00	0.00	0.02	0.13	--	0.12	0.02	0.13
10	0.24	0.24	0.22	0.23	0.00	0.01	0.21	0.23	0.08	--	0.10	0.26
11	0.10	0.06	0.10	0.11	0.01	0.02	0.23	0.04	0.02	0.08	--	0.12
12	0.17	0.14	0.10	0.11	0.01	0.04	0.14	0.13	0.09	0.21	0.10	--

The other articles have less in common with one another, but, by looking at where they stand within the textual network, it is possible to give them the identity of neurophysiology. Highly connected within the neurophysiology hemisphere, words like neuron, potential, culture, rat and hippocampal show which characteristics of neuronal tissues are the focus of the disciplinary advancement and that are about to be linked with carbon nanotubes.

Figure 4 - Textual network. Concepts from Lovat et al. 2005 and its citations. Node size increases in the shortest path betweenness schore and colors are assigned to different groups discovered by a community finding algorithm.



Working only within the network of concepts and eliminating the highest degree centrality nodes which anchored the highest number of concepts, a more divided structure appears. Figure 4 shows the two main families of concepts. On the grey background, the family that belongs to neurophysiology, while externally there are two communities of well interconnected concepts that belong to chemistry. A community finding algorithm highlights them. Neurophysiology concepts are colored in red, chemistry in blue and a set of properties of surface chemistry in green. Only two nodes like brain and nerve are in the chemistry region, while they belong to the organic area, but this depends on their co-occurrence of this little number of abstracts. Adhesion brokers between the concepts related to the chemical compound and those related with its function of support of biological matter. The integration between the two disciplines is sustained by the semiconducting properties of the surface, but mainly by the electrode, which is a common tool across disciplines to measure electrical properties of materials and tissues. Another node laying in between the two ontologies is treatment, but this concept is too generic to be delved further. Other important but peripheral concepts are type of neuronal response – electric spike –, pharmacological treatments, like neurotorphic, or properties of carbon nanotubes, for instance 'single walled' that appears as walled on the right end of the figure.

The network is very compact and the average path distance – how many steps on average it takes for any node to get to any other node – is 1.4452. Despite the higher number of nodes in the next two phases, this measure will diminish slightly also in the following collections by the sole integration of fields and process of convergence.

Phase two: the blend of three disciplines

The integration of a third disciplinary area, engineering, extends the biological information on neurons and its activity especially from an electrical standpoint. It enriches the group with the abstraction of theoretical models, simulation competences and new experimental techniques.

Hierarchical clustering of abstracts that stem from the publication of 2007 in the Journal of Neuroscience does not provide clear cutting points for the definition of coherent clusters, with the exception of the abstract of the paper identified with 14, which is rather dissimilar to others due to the intensive presence of concepts referring to specific techniques.

Figure 5 - Bipartite network: abstracts and concepts stemming from Mazzatenta *et al.* 2005 and its citations.

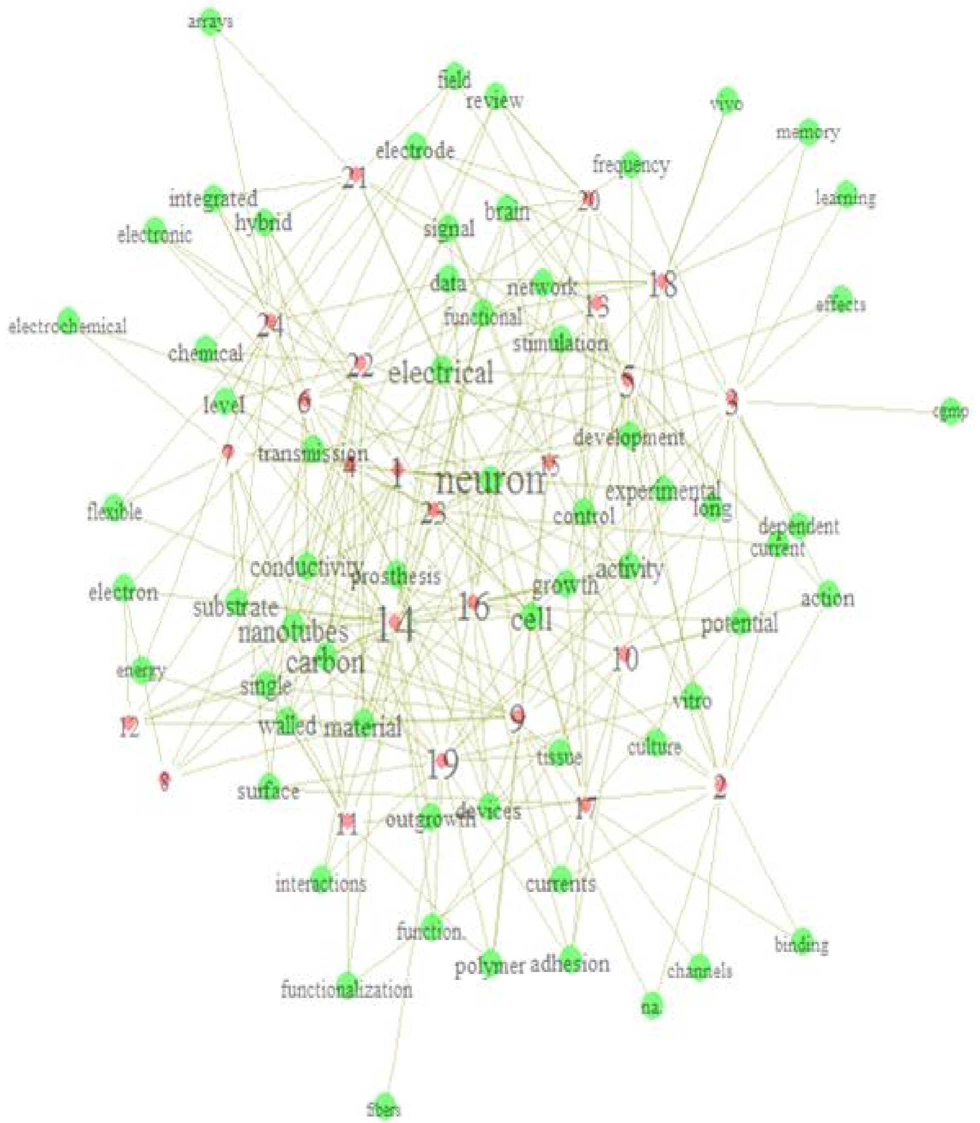
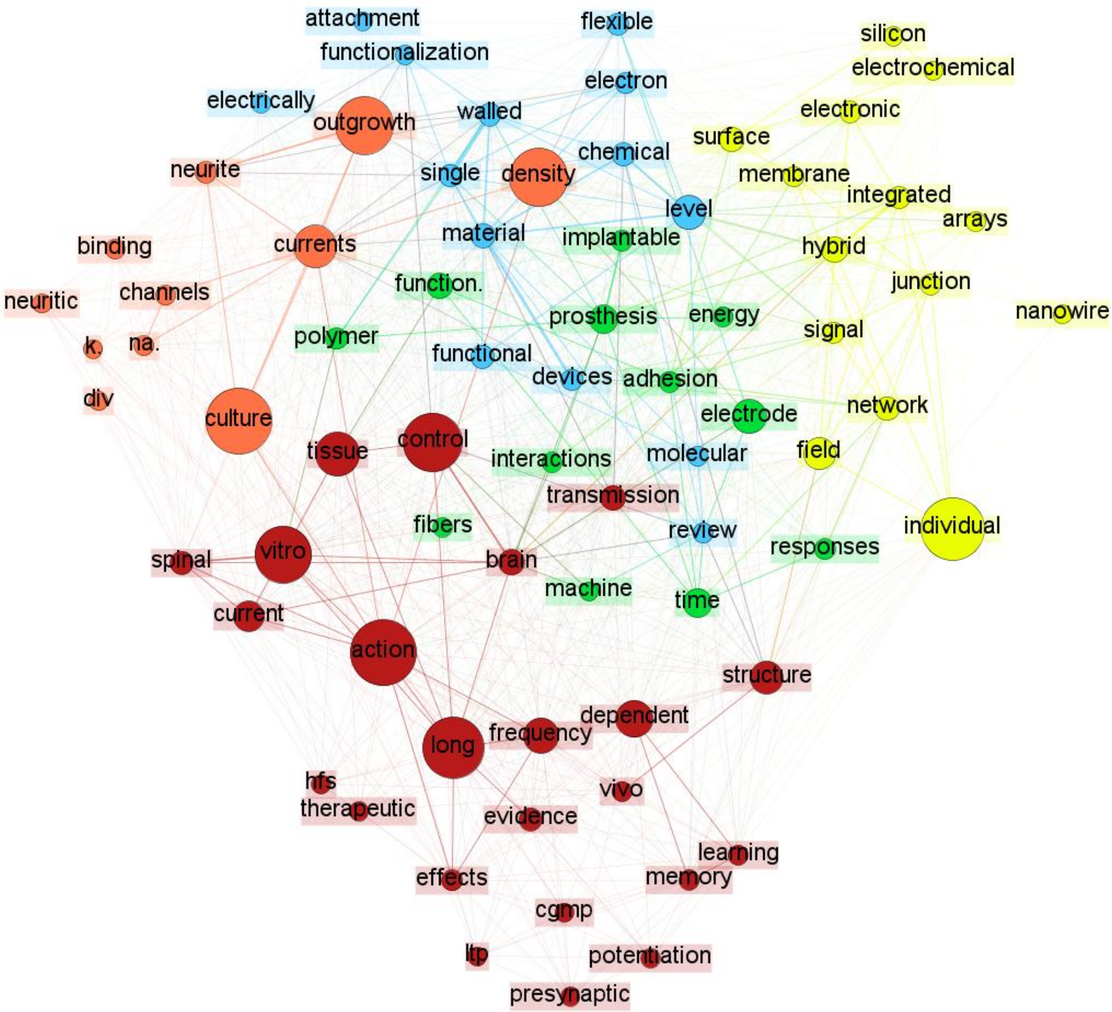


Figure 6 - Textual network. Concepts from Mazzatenta *et al.* 2005 and its citations.



The bipartite network, displayed in Figure 5, where concepts are linked to their origins, abstracts of papers, show an interesting threefold division of articles. In the top area, it is possible to identify engineering related articles with their content. They are specifically focused on the understanding of the electrical activity of neurons at a network level of analysis, a novelty for the project. New techniques, such arrays of electrodes become part of the knowledge sets. Notwithstanding their peripheral position, memory and learning point out what is the core of attention of the neural analysis at the bioengineering level. In the middle, a wedge shaped area characterizes concepts typical to neurophysiology. On the lower left area, the chemistry concepts.

At this stage, the amount of concepts is roughly equally split across disciplines, and the integration – bridging – between disciplinary areas does not rely anymore on a couple of critical concepts.

The textual network shown in Figure 6 confirms the idea that concepts are better meshed and the network is more compact. Chemistry concepts does not appear to be easily detachable from the others, because the interaction with other other-disciplinary terms is stronger. That means that concepts coming from diverse disciplinary co-occur in the abstracts used for the second article. Community finding algorithm identifies the blue and green partitions that overlap chemistry concepts, although some other concept like nanowire or electrochemical are in the yellow group, meshed together with bio-engineering words on the top-right part of the figure. Single walled (nanotubes, cut off due to the high degree centrality) are in the middle of the network. On the left, in orange, we identify concepts specific to neurophysiology such as sodium, potassium, ‘channels’, ‘culture’. Engineering specific terms are on the bottom such as ‘frequency’, ‘learning’, as well as on the yellow community of concepts, where ‘network’ and ‘signal’ lay. Action potential, the electrical signal fired by neurons, and it is the core of study of neurophysiology and engineering seems to bridge the two otherwise non separated ontologies. Indicators of betweeenness in Table 3 confirm the conceptual bridge of action potential, that is the electrical signal of neurons.

It is surprising that by adding a new knowledge base, bio-engineering, which carries its abstract and lexicon, there is not a negative effect on the integration between fields as the average path length has slightly decreased to 1.4147. This confirms that the lexicon of bio-engineers partly overlaps with that of neurophysiologists.

Phase three: the applied nature of chemistry

The last set of abstracts under scrutiny stems from the last experimental paper of these scientists within this collaboration we considered. The knowledge sets are still three and the hierarchical clustering shows clearly these three groups, see Figure 2, plot on the right. Two of them are rather close to each other, but they are very distant from the third.

Figure 7 - Bipartite network: abstracts and concepts stemming from Cellot *et al.* 2005 and its citations.

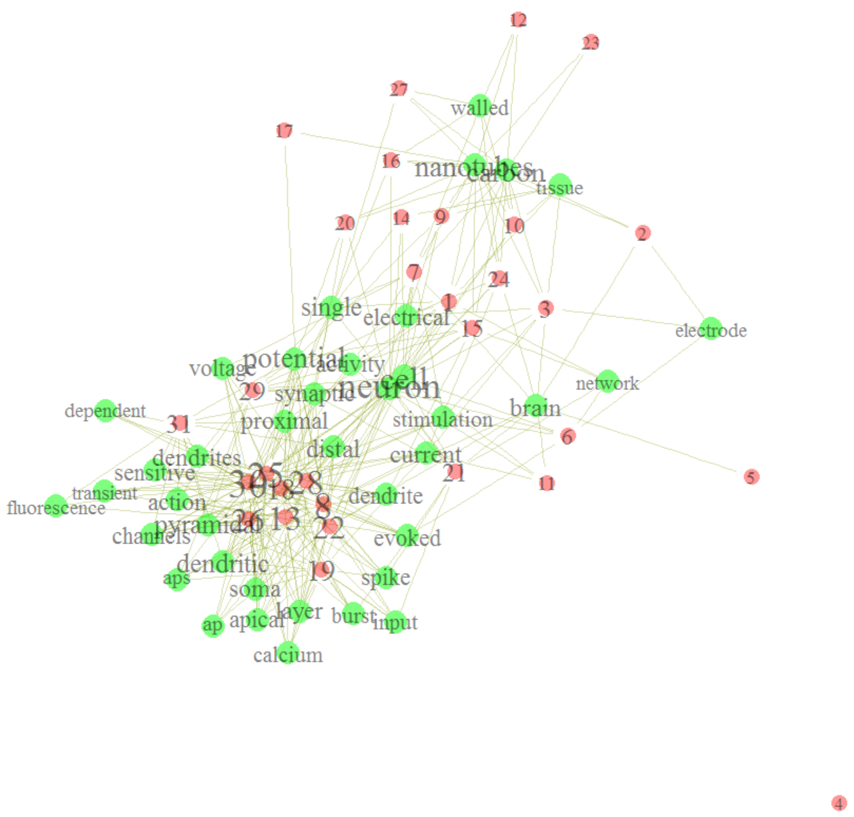
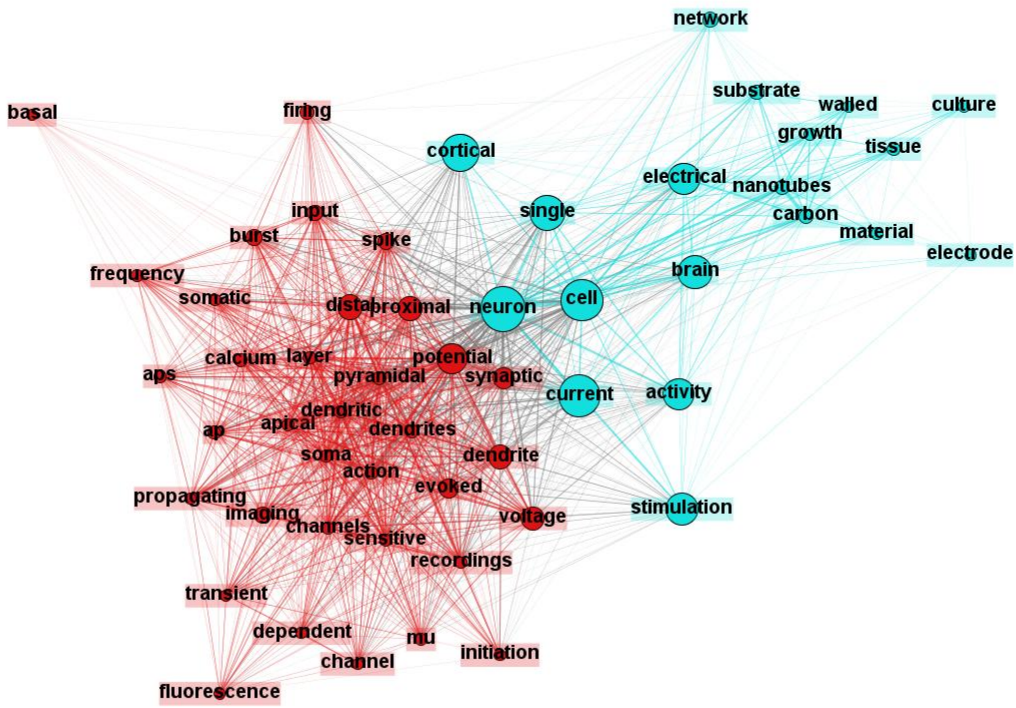


Figure 8 - Textual network. Concepts from Cellet *et al.* 2005 and its citations.



A visual combination of words and articles, see Figure 7, shows a central core of concepts that divides chemistry words such as carbon, nanotubes and walled that lay on the top from biology related concepts like dendrites, soma, calcium or electrically related to them, ap (action potential), evoked burst. Interestingly, the interface between the chemical ontology and the other two is no longer revealed by tools of measure such as an electrode or by fuzzy concepts of transmission and treatment, but rather by the electrical properties of neuronal tissues are the new points of connection. Many articles mention neurons and carbon nanotubes together in the abstracts and this show that the association has been incorporated in the scientific literature. Chemistry concepts are marginalized at the periphery of the network. Chemistry becomes functional to other disciplinary worlds in the understanding of the electrical activity of neurons interacting with the carbon substrate.

This is confirmed by the analysis of betweenness (see § Measures of betweenness) along with the visual representation of the textual network in Figure 8. In this case, two communities of concepts are clearly identified. Chemistry stays on the right, in blue, and engineering and neurophysiology are in red. New concepts bridge across disciplines. Concepts at the boundary are imaging that recall the importance of microscopy as instrument of inquiry and for the communication across disciplinary fields especially in the world of Nanotechnology, cell, neuron, dendrites. This analysis confirms that the new dimension of the joint collaboration and generally of the research in the field enquire the characteristics of the neuron at the level of the electrical properties measured along its elongations.

Despite the peripheral position of chemistry, the overall path distance has shortened to 1.3520. This show a higher integration of fields.

Measures of betweenness

In this section, we test different measures of betweenness to capture cross-boundary concepts in the textual networks. In Table 3, we show the first 12 results according to each measure computed for each collection of abstracts.

Shortest-path betweenness (Freeman 1979) measures for any node how many geodesic paths between any two nodes go throw it. In the first article, treatment and single bridge words in the fields of neurophysiology, identified red through the modularity algorithm, see Figure 3, with concepts that belongs to chemistry. Treatment and single, however, are ambiguous. Treatment may have different meanings. In neurophysiology, it may indicate pharmacological modifications of a substance and it is a typical word used in the experimental settings. In chemistry, its meaning broadly represents the processes of manipulation of a substance. Single seems a common word, but it has not been excluded because it may grasp a particular type of carbon nanotubes (single walled, instead of multi walled). More interesting results are electrode, amplitude and adhesion. Adhesion is a necessary and critical concept in the work of Lovat and colleagues' as it transforms carbon nanotubes from a purely chemistry matter to a compound that may be coupled with a biological substrate. Adhesion to glass is a property of carbon nanotubes that enable scientists to perform the experiment. Amplitude has high betweenness because it is a common concept in neurophysiology. It does not bridge disciplines. Unlike, electrode is the physical object at the boundary (Star 1989), as it inhabits more ontologies. It is used to measure the electrical activity on neurons and it ties the set of nodes of surface chemistry with the set of concepts of neurophysiology. In the second stage of the collaboration, the network is tightly connected. Action (potential) – the electric signal sent by neurons – is a common concept between the newly entered field of bio-engineering and neurophysiology, whereas outgrowth bridges chemistry and neurophysiology showing that the literature has focused on the neuronal growth on layers of functionalized carbon nanotubes. In the last collection of articles, chemistry words look more distant from a core of words that meshes concepts of engineering and neurophysiology. SPB scores identify those nodes who sit between these two sets.

Flow betweenness (Freeman et al. 1991) depends on the weight of the edges of paths. In this case, in the frequency of co-occurrence within documents. If a node is within high capacity ties, its FB score will be higher than those within the same network of edges with lower capacity. Ties around carbon and nanotubes have larger weights comparatively with others, therefore the FB scores of carbon and nanotubes are high. Surface and electrode are the two concepts linking the ontologies. In the second collection of abstracts, FB scores identify prosthesis that is not a concept useful for the joint research, but it is one of the potential applications for which the carbon nanotubes are prepared (Tasis et al. 2006). In the last phase of the collaboration, FB scores yield similar results to those of SPB. Nodes that lay in between fields emerge in the highest position, notwithstanding the bias towards higher capacity edges.

Power centrality (Bonacich 1987) scores embeds the degree of neighbor nodes. If the neighbors of node *i* have large degree, then *i* will have a high PC score. The larger the β, the larger the influence of neighbors. For the small average distance in the three networks, we chose a low value of β equal to 0.2; higher values point to nodes in the periphery. We experimented also values of 0.5 and 1. In the first article, gap and acid are linked with high degree nodes, but they are not meaningful. Electrode is the first boundary word to appear in the list, adhesion comes after. Nerve is an example of how this measure may misfire, because it is in the periphery and it does not bridge ontologies. In the second set of articles, the first results of the PC index are in the periphery of the network. Unlike, in the third set of articles PC scores highlight nodes in between fields, this is due to the nature of the connection between ontologies that occurs through nodes that are tied with high degree nodes. However, PC provides poor results, despite the low value of β.

Random walk betweenness (Newman 2005) measures the probability for a random walk between any two nodes to pass through any given node. In the first phase, transmission is the only new interesting result, as it brokers between carbon nanotubes and more biological concepts represented by words like excitatory and synaptic. In the second stage of the collaboration, RWB generates a ranking similar to that of SPB, with a few differences: brain is the first results. Brain connects the concepts of single walled (carbon nanotubes) with the organic world. In the last phase, RWB scores show dendrites – the neuronal elongations – in the first results. They are the new focus both for neurophysiology and bio-engineering in the understanding of the interaction with carbon nanotubes. It is on the dendrites that most of the attention of simulations and electrical recording resides.

In summary, no measure outperformed all others in the search for words at the boundary. Nonetheless, the best results are given by the SPB and RWB, with a bias by the latter to favor nodes that sits along high capacity ties. FB is also a good measure, but it seems to be so for the assumption that a word is connected to all others if belongs to the same abstract. This creates large chunks of words, and, in turn, compact networks. It may not be the case in more sparse networks. The poorest index, however not meant to be a measure of betweenness, but of influence in social networks, is the PC. Even at low values of β, this index is biased by the proximity of nodes with high degree that in turn have high degree nodes attached. In our cases, this cascade fashion favors vertices that sit on the periphery, instead of nodes that broker across disciplines.

Table 3 - Measures of centrality. Shortest-path betweenness (SPB), Flow betweenness, Power centrality, and Random-walk betweenness (RWB)

Lovat *et al.* 2005 - Nano Letters
without highest centrality tokens

SPB		Flow Betweenness		PC, β= 0.2		RWB	
treatment	0.058	amplitude	0.039	gap	0.037	treatment	0.117
single	0.045	excitatory	0.039	acid	0.034	amplitude	0.105
electrode	0.042	treatment	0.036	contrast	0.033	excitatory	0.105
amplitude	0.035	nanotubes	0.036	electrode	0.031	contrast	0.100
excitatory	0.035	carbon	0.036	amplitude	0.030	nanotubes	0.096
contrast	0.033	surface	0.033	excitatory	0.030	carbon	0.096
acid	0.031	contrast	0.033	grown	0.030	acid	0.093
gap	0.029	acid	0.032	function.	0.028	electrode	0.090
network	0.026	electrode	0.031	nerve	0.028	transmission	0.090
silicon	0.022	couple	0.030	postsynaptic	0.027	surface	0.088
surface	0.022	extracellular	0.030	tissue	0.026	network	0.087
adhesion	0.020	postsynaptic	0.029	adhesion	0.025	single	0.084

Mazzatenta *et al.* 2007 - Journal of Neuroscience
without highest centrality tokens

SPB		Flow Betweenness		PC, β= 0.2		RWB	
action	0.040	action	0.028	electrode	0.070	action	0.077
culture	0.034	long	0.024	flexible	0.041	culture	0.067
individual	0.026	vitro	0.023	silicon	0.039	long	0.066
long	0.025	culture	0.022	electrochemical	0.039	currents	0.063
control	0.023	currents	0.022	learning	0.033	vitro	0.062
outgrowth	0.022	level	0.022	memory	0.033	outgrowth	0.062
vitro	0.022	outgrowth	0.022	network	0.033	level	0.061
density	0.021	hybrid	0.021	vivo	0.032	control	0.061
currents	0.019	control	0.021	surface	0.032	material	0.058
level	0.018	material	0.021	functional	0.031	individual	0.058
tissue	0.016	prosthesis	0.020	outgrowth	0.031	brain	0.058
dependent	0.016	density	0.020	control	0.029	density	0.057

Cellot *et al.* 2009 - Nature Nanotechnology
all tokens - degree in parenthesis

SPB		Flow Betweenness		PC, β= 0.2		RWB	
neuron (634)	0.079	neuron (634)	0.056	brain (198)	0.026	neuron (634)	0.121
cell (456)	0.040	cell (456)	0.041	electrical (196)	0.025	cell (456)	0.097
potential (482)	0.018	potential (482)	0.036	network (78)	0.023	potential (482)	0.074
single (280)	0.014	dendritic (444)	0.031	current (318)	0.023	single (280)	0.063
current (318)	0.013	pyramidal (438)	0.030	cell (456)	0.023	current (318)	0.061
brain (198)	0.008	distal (374)	0.026	activity (220)	0.023	dendritic (444)	0.059
cortical (218)	0.008	soma (366)	0.025	cortical (218)	0.022	nanotubes (176)	0.059
electrical (196)	0.007	action (360)	0.025	potential (482)	0.022	carbon (174)	0.059
activity (220)	0.007	current (318)	0.025	nanotubes (176)	0.022	pyramidal (438)	0.058
stimulation (188)	0.007	layer (358)	0.025	carbon (174)	0.022	electrical (196)	0.058
distal (374)	0.006	single (280)	0.024	dendrite (320)	0.022	brain (198)	0.056
proximal (332)	0.005	dendrites (358)	0.024	single (280)	0.022	distal (374)	0.056

DISCUSSION

The combination of textual analysis with network theory has been used to visualize disciplinary boundaries in a cross-disciplinary collaboration of scientists. This is an important step to cast light into the conceptual dynamics of boundaries, disciplinary relations and cross-disciplinary and more generally cross-boundary collaborations. By looking at the abstracts of three papers published by a cross-disciplinary scientific collaboration, along with abstracts of their cited articles, we see that boundaries are dynamic (Abbott 1995), relations between disciplines evolve, and ontologies are subject to changes, adaptation and partial blend.

In the three timeframes analyzed within the collaboration, disciplinary boundaries have moved not just for being distinguished from the external world, as addressed by most literature on scientific boundaries (Lamont, & Molnar 2002), but from the effort of scientists to combine knowledge sets in the joint exploration. This work shows that disciplines have proper characteristics, such as methods and targets, which are visible in the concepts that appear around specific materials. For example, functionalization, adhesion lay close to carbon nanotubes and are part of the jargon of chemistry. Such disciplines, however, can be differently integrated over time.

Facilitating cross-boundary communication and coordination by means of tools is possible when they make already sense in both worlds(Levina and Vaast, 2005), and therefore when they have already an established network of meanings. Similar to the idea of boundary objects (Star 1989; Star & Griesemer 1989) that enable the organization of work across boundaries, conceptual bridges allow the coordination and the collaboration of different worlds and, thereby, they are boundary spanners. Examples of such boundary spanning objects are the electrode that bridges neurophysiology and chemistry in the first part of the collaboration or the dendritic mesh in the latter part. We thus provide evidence that conceptual bridges are not static. On the contrary, they change over time according to the project – goal – and the combination of knowledge bases. This technique shows that they can be visualized and they have higher scores of betweenness centrality as they are points of connection between ontologies. Moreover, being able to see (and foresee) which concepts-objects bridge ontologies could facilitate cross-boundary communication and coordination, as we know that when objects are forced between boundaries they not necessarily succeed in their purpose of boundary objects (Levina and Vaast 2005).

Our results show that in the first part of the collaboration, the two ontologies are mostly separated, with the exception of few bridges that establish a connection between disciplinary worlds. In the second part of the collaboration, with a third discipline involved, it is visible how each discipline has a clear conceptual territory and the joint structure is an ordered aggregation of three different conceptual spaces, in which neurophysiology stands as interface between the other two disciplines, and the integration occurs through several – and not anymore few – conceptual bridges. In the last phase, concepts of neurophysiology and bio-engineering could not be partitioned differently, since they are embedded in abstracts with overlapping concepts, whereas the distance with respect to chemistry related concepts augments. Moreover, chemistry articles cited tend to adopt more biological words like neurons and cell, thus indicating a changing nature of the cited articles of chemistry that seem to be more contextually applied.

Following the idea that understanding is grounded on common beliefs and knowledge (Clark, 1996), this method sheds light on the points of contact – the common ground – that could enable cross-cultural communication. Indeed, this common ground, as it inhabits multiple worlds, is at the boundary and brokers. Still, it may leave open ambiguous interpretations, but they could be traced back by looking at concepts linked to the common one that inhabit different ontologies. When objects they could provide the tangible definitions for contextualization necessary to solve misunderstanding (Bechky, 2003).

Usually when looking at cross-boundary relations, boundaries are considered as fixed, because it is hard to gather data to look at its nature. In a context similar to this study's, Cummings' and

Kiesler's (2005) analysis on scientific multidisciplinary ventures and their coordination mechanisms treated disciplines as static, ignoring what occurs while collaboration takes place. By looking at disciplinary conceptual spaces and their interaction, management theory can first operationalize disciplinary distance taking into account concepts. Then, theorists can make predictions on the types of mechanisms necessary to link disciplines that have different distance: from few or no points of contact to many. For this reason, we believe that this is not a mere visualization tool that aids researchers to support their qualitative insights, but, by combining the power of network theory with its statistical methods, it has heuristic value as it may help theorizing about boundaries and give the tools to test hypotheses.

CONCLUSIONS

In this article we show a method to visualize cross-boundary concepts in multidisciplinary scientific research. Using the abstract of a scientific article and those of its citations, we collect texts from multiple disciplines from which we built a document term matrix. An adjacency matrix of words is created and links appear when words co-occur in the same document. Therefore such a textual network shows connections between words of different documents.

This allows study a multidisciplinary scientific context in its dynamics and see how disciplines change their relations over time and what are the concepts that create bridges. We provide a first attempt to give a visual form to the study of boundaries and we can import mathematical measures of centrality to see which best captures concepts that inhabit multiple disciplinary spaces. Our tests show that shortest path betweenness (Freeman 1979) and random walk betweenness (Newman 2005) perform better than flow betweenness (Freeman et al. 1991) and Power centrality (Bonacich 1987) in such networks. However, further analysis should be devoted to understand which measure captures such critical concepts in a larger number of textual networks.

We show that concepts that bridges disciplinary knowledge domains may change over time and due to the aim of the joint-research project. A theory of boundaries in collaboration across distinct knowledge field may emerge in the future, and this could be an initial step towards a new type of theories.

We believe that this method has strengths as well as limitations. One of the limitations is due to the boundary conditions of the text capturing assumption. Although relations is established between any two words in an abstract and different abstracts are linked by same words, we cannot assume that the meaning concept underlying that word is shared across abstracts. Yet, the same concept draws on different set of relations, analytical tools, epistemic foundations that are not necessarily shared. For instance, behind the concept of network analysis there are different meanings and different techniques are used to study it according to the epistemic community that adopts it. Social network scholars maps relations between actors or firms and produce graphs to treat the concept mathematically; economic sociologists may highlight the social, cognitive and economic conditions that enable the formation of a interconnected set of firms, not necessarily attracted by the link between firms. For this reason, the technique is not mature yet to be unsupervised, because it could provide ambiguous or even misleading results if applied to any set of abstracts. Moreover, the supervision of a researcher is also needed to understand the data in order to distinguish fields. But in this respect, we believe that a further step towards automatic detection of disciplinary jargons in corpora of texts could be done through the creation of disciplinary. A second important limitation linked to the text capturing assumption is that it impedes the generalization of such method in contexts where texts come from different sources, such as interviews or not documents not as simplified as abstracts. The same collections of texts may feed algorithms that measure differently the semantics of words (e.g., Lund and Burgess 1996). This would produce different adjacency matrices that, in turn, may generate different boundary relations across disciplinary concepts.

A strength of this method, instead, is that it is possible to apply it to other languages than English, to other forms of texts which still must be small and coherent for the assumption that any word is related to any other within the text.

We believe that with such a refined method may aid qualitative researchers in resolving ambiguity of words and constructs and showing different viewpoints of interviewees if adopted on proper chunks of texts. Providing a mathematical tool to read words, it could help the passage between a theory-building type of approach to a theory-testing phase. For example, at the light of social network studies (e.g., Ibarra and Andrews, 1993; Padgett and Ansell 1993), actors who are in between tend to have more power and control, while actors who withhold a peripheral position could be more easily substituted or have less power on the control of relevant information. Our results show that chemistry becomes functional to the other core of the experimental study which is maneuvered by neurophysiologists and bio-engineers. We do not know whether such peripheral conceptual position of chemistry is contingent on the experiment itself or due to an ongoing marginalization and our data is too little to resolve this doubt. As such and despite limitations in case of polysemy due to presence of epistemic communities, this study presents a method that could be visually helpful and statistically robust to refine theories of boundaries.

REFERENCES

Abbott, A. 1995. "Things of Boundaries". ***Social Research***, 62: 857-882.

Acedo, F.J., Barroso, C., Casanueva, C., & Galán, J. L. 2006. "Co-Authorship in Management and Organizational Studies: An Empirical and Network Analysis". ***Journal of Management Studies***, 43: 957-983.

Amabile, T. M. 1988. "A Model of Creativity and Innovation in Organizations". ***Research in Organizational Behavior***, 10: 123-167.

Amine, A., Elberrichi, Z., & Simonet, M. 2010. "Evaluation of Text Clustering Methods using WordNet". ***The International Arab Journal of Information Technology***, 7, 349-356.

Bauer, H. H. 1990. "Barriers against interdisciplinarity: Implications for studies of science, technology, and society". ***Science, Technology & Human Values***, 15, 105–119.

Bechky, B. A. 2003. "Sharing Meaning Across Occupational Communities: The Transformation of Understanding on a Product Floor". ***Organization Science***, 14: 312-330.

Bhattacharya, S., & Basu, P. K. 1998. Mapping a research area at the micro level using co-word analysis. ***Scientometrics***, 43: 359-372.

Blondel, V. D., J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, 2008. "Fast Unfolding of Communities in Large Networks". ***Journal of Statistical Mechanics***, P10008.

Bonacich, P. 1987. "Power and Centrality: A Family of Measures". ***American Journal of Sociology***, 92: 1170-1982.

Borgatti S. P. 2005. "Centrality and network flow". ***Social Networks***, 27: 55-71.

Bowker, G, & Star, S. L. 1999. ***Sorting Things Out: Classification and Its Consequences***. Cambridge, MA: MIT Press.

Callon, M., Courtial, J. P., & Laville, F. 1991. "Co-word analysis as a tool for describing the network of interactions between basic and technological research: The case of polymer chemistry". ***Scientometrics***, 22: 153- 205.

Carlile, P. R. 2004. "Transferring, Translating, and Transforming: An Integrative Framework for Managing Knowledge Across Boundaries". ***Organization Science***, 15: 555-568.

Cenci, C., Pozzi, E., Borsacchi, M. 2010. "Autopsia Semantica di un Corpo Mistico: la Morte di Michael Jackson nella Stampa Italiana e su Youtube". ***Proceedings of JADT 2010***, 10th International Conference on Statistical Analysis of Textual Data, Rome, Italy, 607-617.

Caldas, C. H., & Soibelman, L. 2003. "Automatic hierarchical document classification for construction management information systems". ***Automation in Construction***, 12: 395-406.

Clark, H. 1996. ***Using Language***, Cambridge: Cambridge University Press.

Csardi, G., Nepusz, T. 2006. "The igraph software package for complex network research". ***InterJournal, Complex Systems***. 1695.

Cummings, JN, & Kiesler, S. 2005. "Collaborative Research Across Disciplinary and Organizational Boundaries". ***Social Studies of Science***, 35: 703-722.

Degenne, A. , & Forse, M. 1999. ***Introducing Social Networks***, London: SAGE Publications Ltd.

Ding, Y., Chowdhury, G. D. , Foo, S. 2001. "Bibliometric cartography of information retrieval research by using co-word analysis". ***Information Processing and Management***, 37, 817-842.

Epstein CF. 1992. "Tinker-bells and pinups: the construction and reconstruction of gender boundaries at work". In Lamont & Fournier (Eds.) ***Cultivating Differences: Symbolic Boundaries and the Making of Inequality***. Chicago: Univ. Chicago Press 1992, pp. 232–56.

Eto, H. 2003. "Interdisciplinary information input and output of nano-technology project". ***Scientometrics***, 58:5–33

Faraj, S., & Sproull, L. 2000. "Coordinating expertise in software development teams". ***Management Science***,46: 1554-1568.

Ford, L.R., & Fulkerson, D.R. 1962. ***Flows in Networks***. Princeton, NJ: Princeton University

Freeman, L.C. 1977. "A set of measures of centrality based upon betweenness". ***Sociometry***, 40: 35-41.

Freeman, L.C. 1979. "Centrality in Social Networks: conceptual clarification". ***Social Networks***, 1: 215-239.

Freeman, L. C., Borgatti, S. P., & White, D. R. 1991. "Centrality in valued graphs: A measure of betweenness based on network flow". ***Social Networks***, 13: 141-154.

Gianvecchio S, Xie M, Wu Z, and Wang H. 2008. "Measurement and classification of humans and bots in internet chat". ***SS' 08: Proceedings of the 17th Conference on Security Symposium***, pp. 155-169.

Gieryn, T. F. 1999. ***Cultural Boundaries of Science: Credibility On the Line***. Chicago, IL: Univ. Chicago Press.

Goh, K.-I., Oh, E., Kahng, B., Kim, D., 2003. "Betweenness centrality correlation in social networks". ***Physical Review E***, 67, 017101.

Hargadon, A. B., & Sutton, R. I. 1997. "Technology brokering and innovation in a product development firm". ***Administrative Science Quarterly***, 42: 716-749.

Hirst, G., Riabinin, Y, Graham, J. 2010. "Party Status as a Confound in the Automatic Classification of Political Speech by Ideology". Proceedings of JADT 2010, 10th International Conference on Statistical Analysis of Textual Data, Rome, Italy, 731-742.

Hsiao, R-L. , Tsai, D-H., & Lee, C. F. 2012. "Collaborative Knowing: the Adaptive Nature of Cross-Boundary Spanning". ***Journal of Management Studies***, 49: 463-491.

Ibarra, H., & Andrews, S. B. 1993. "Power, Social Influence, and Sense Making: Effects of Network Centrality and Proximity on Employee Perceptions". ***Administrative Science Quarterly***, 38: 277-303.

Joachim, C., & Plevert, L. 2009. ***Nanosciences. The Invisible Revolution***. World Scientific Publishing Company.

Lamont, M., Molnar, V. 2002. "The Study of Boundaries in the Social Sciences". ***Annual Review of Sociology***, 28: 167-195.

Leonardi, P. M. 2011. "Innovation Blindness: Culture, Frames,and Cross-Boundary Problem Construction in the Development of New Technology Concepts". ***Organization Science***, 22: 347-369.

Levina, N., & Vaast, E. 2005. "The Emergence of Boundary Spanning Competence in Practice: Implications for Implementation and Use of Information Systems". ***MIS Quarterly***, 29: 335-363.

Leydesdorff, L. 2007. "Betweenness Centrality as an Indicator of the Interdisciplinarity of Scientific Journals". ***Journal of the Maerican Society for Information Science and Technology***, 58: 1303-1319.

Looze, M. D., & Lemarie, J. 1997. "Corpus relevance through co-word analysis: An application to plant proteins." ***Scientometrics***, 39(3), 267-280

Lund, K. & Burgess, C. 1996. "Producing high-dimensional semantic spaces from lexical co-occurrence". ***Behaviour Research Methods, Instruments, & Computers***, 28: 203-208.

Markovsky, Barry. 1998. “Social Network Conceptions of Solidarity.” Pp. 343–72 in ***The Problem of Solidarity: Theories and Models***, edited by Patrick Doreian and Thomas Fararo. Amsterdam, The Netherlands: Gordon and Breach

Moody, J., & White, D. R. 2003. “Social Cohesion and Embeddedness: A Hierarchical Conception of Social Groups”. ***American Sociological Review***, 68: 103-127.

Moody, J. 2004. “The Structure of Social Science Collaboration Network: Disciplinary Cohesion from 1963 to 1999”. ***American Sociological Review***, 69: 213-238.

Murtagh, F., Ganz, A., & McKie, S. 2009. “The structure of narrative: the case of film scripts”. ***Pattern Recognition***, 42, 302-312.

Newman, M. E. J. 2001. “The Structure of Scientific Collaboration Networks”. ***Proc Natl Acad USA,*** 98: 404-409.

Newman, M. E. J. 2004. “Coauthorship Networks and Patterns of Scientific Collaboration”. ***Proc Natl Acad USA***. 101, Suppl 1: 5200-5205.

Newman, M. E. J. 2005. “A measure of betweenness centrality based on random walks”. ***Social Networks***, 27: 39-54.

Newman, M. E. J. 2004. “Analysis of Weighted Networks”. ***Physical Review E***, 70: 056131.

Newman, M. E. J. 2010. ***Nerworks: An Introduction***. Oxford, UK: Oxford University Press.

OECD 2010. ***Measuring Innovation: a new perspective***, OECD Publishing.

Oswick, C. Fleming, P., & Hanlon, G. 2011. “From Borrowing to Blending: Rethinking the Process of Organizational Theory-Building”. ***Academy of Management Review***, 36: 317-337.

Padgett, J. F. & Ansell, C. K. 1993. “Robust Action and the Rise of the Medici, 1400-1434”. ***American Journal of Sociology***, 98: 1259-1319.

Parkin,F. 1974. “Strategies of Closure in Class Formation”. In ***The Social Analysis of Class Structure***, 1-18.

Perper, T. 1989. “The loss of innovation: Peer review in multi- and interdisciplinary research”. ***Issues in Integrative Studies***, 7, 21–56.

Pierce, S. J., 1999. “Boundary Crossing in Research Literatures as a Means of Interdisciplinary Information Transfer”. ***Journal of the American Society for Information Science***, 50: 271-279.

Porter, A. L. & Youtie, J. 2009. “How interdisciplinary is nanotechnology?”. ***Journal of Nanoparticle Research***, 11: 1023-1041.

Roco, M. C., and W. S. Bainbridge, eds., ***Converging Technologies for Improving Human Performance***, NSF-DOC Report.

Rose, S., Engel, D., Cramer, N. & Cowley, W. 2010. “Automatic keyword extraction from individual documents”, in Berry M. W. & Kogan J. (Eds.), ***Text Mining: Applications and Theory***, pp: 3-20, Wiley: Chichster: UK.

Schummer, J. 2004. “Multidisciplinarity, Interdisciplinarity, and Patterns of Research Collaboration in Nanoscience and Nanotechnology”. ***Scientometrics***, 59: 425-465.

Shapira, P., Youtie, J., Porter, A. L. 2010. “The Emergence of Social Science Research on Nanotechnology”. ***Scientometrics***, 85: 595-611.

Smith, A. 1776. ***An Inquiry into the Nature and Causes of the Weath of Nations***. London, UK: Strahan and Cadell.

Small, H. 1972. “Bibliometric indicators of the development of nuclear physics”. ***The Physical Review***, 1927–1934.

Star, S. L., & Griesemer, J. R. 1989. “Institutional Ecology, ‘Translations’ and Boundary Objects: Amateurs and Professionals in Berkeley’s Museum of Vertebrate Zoology, 1907-39.” ***Social Studies of Science***, 19, 387-420.

Star, S. L. 1989.” The Structure of Ill-Structured Solutions: Boundary Objects and Heterogeneous Distributed Problem Solving.” In Gasser, L. & Huns, M. N. (Eds.) ***Distributed Artificial Intelligence***, vol. 2, San Francisco, CA: Morgan Kauffman Publishers, Inc. 37-54.

Tasis, D., Tagmatarchis, N., Bianco, A. & Prato, M. 2006. “Chemistry of Carbon Nanotubes.” ***Chem. Rev.***, 106: 1105-1136.

The royal society & The royal Academy of Engineering 2004 – ***Nanoscience and nanotechnologies: opportunities and uncertainties***. Available online at http://www.nanotec.org.uk/report/Nano%20report%202004%20fin.pdf

Van Rijsoever, F.J. & Hessels, L.K. 2011. “Factors associated with disciplinary and interdisciplinary research collaboration.” ***Research Policy***, 40: 463-472.

3. STANDING ON GIANTS' SHOULDERS. HOW NETWORKS' STRUCTURE AND THEMATIC INNOVATION AFFECT THE SCIENTIFIC IMPACT IN THE VULNERABILITY LITERATURE

Claudio Biscaro

Ca Foscari University of Venice

Carlo Giupponi

Ca Foscari University of Venice

ABSTRACT

This work analyzes what affects the citations received by scientific articles in the literature of climate change vulnerability. Specifically, we measure the effects of two different networks generated by the position of co-authors in the field, articles within the literature, along with the innovativeness of the articles' content. By including variables that are proper to the knowledge embedded in the article, we expand prior research which limited its focus to the position of co-authors. By means of a machine learning algorithm applied to natural language, we assign to each articles one or more topics and define two types of innovation: innovative are those articles which introduces a new topic in the literature or that provide new combinations. Then for each article we compute centrality measures (*degree centrality*, *betweenneess centrality*, and *closeness centrality*) from the co-authorship and the bibliographic coupling network to see how they influence the citations received by articles in first two years after the year of publication. We find that new topic combinations and generations have no effect on the *citation count*, whereas author's degree positively impacts citations, while betweenness has a negative effect. Moreover, we find that articles that draw on fragmented literature tend to be cited more.

Keywords: knowledge combination, bibliographic coupling, co-authorship network, topic modeling, scientific impact.

INTRODUCTION

Dicebat Bernardus Carnotensis nos esse quasi nanos gigantium humeris insidentes, ut possimus plura eis et remotiora videre, non utique proprii visus acumine, aut eminentia corporis, sed quia in altum subvenimur et extollimur magnitudine gigantean. [***Metalogicon III, 4***, John of Salisbury]

Bernard of Chartres used to say that we are like dwarfs on the shoulders of giants, so that we can see more and more distant things than them, nor for our own sharp sight, neither for the superiority of the body, but because we are carried high and elevated by the stature of giants.

The inspiring portrait of scientist building on the work of other scientists seems to be self-explanatory and convincing, but it brings about two deep and important puzzles. Where exactly is it better to stand in order to see more and more distant things? And what knowledge should we use?

Along the same idea of a scientific evolution, Polanyi (2000) noticed that scientists adjust their efforts on the basis of "the hitherto achieved results of the others" (p. 2) thus creating a continuum of different works that is in a continuous progress. Such an evolving patchwork can be represented through networks that, on the one hand, show how collaborations among scientists create bridge scientific fields (Newman, 2001; Sun, Kaur, Milojevic, Flammini, & Menczer, 2013) and, on the other, use articles as nodes to look at the change of the main sources of knowledge over time (Chen, 2004; Janssen, 2007; Janssen, Schoon, Ke, & Borner, 2006; Sun et al., 2013). Yet, although much work has been done to highlight the evolving nature of science and scholarly collaborations, it is not clear how to disentangle what types of contributions have a greater impact.

A traditional measure of impact is the citations that an article receive. They are given by authors who recognize an article's value and decide to draw on its ideas, or eventually to challenge them. Citations are embodied in the determinants of scholarly career paths set by universities, who privilege scientists who publish on top-tier journals (which become such due to the high number of citations received by the articles included). Therefore, citations are the primary or, at times, the exclusive measure of scientific impact and the kernel of the attention of scholars work who, through bibliometric studies, try to understand what makes articles relevant.

This movement goes back to the early work of Garfield (1955) that establishes that the success of a journal, and of the ideas therein contained, can be quantified by their number of citations. Since then, citations have also been used to describe the history and the evolution of science through maps, as in the works of Henry Small in the 70s and 80s, and their properties gathered the attention of many scholars who studied their distribution over the years (Redner, 1998), the advantage of the first publication in a field (Newman, 2009), the bandwagon effect of accolades and world-class recognitions on the rewarded scientist's prior works (Mazloumian, Eom, Helbing, Lozano, & Fortunato, 2011). They also have been related to the network position of co-authors and it has been found that some centrality measures significantly correlate with the article citation count (Yan & Ding, 2009), particularly betweenness centrality (Freeman, 1979) – that measures how many shortest paths goes through a node – is the highest correlated metric.

The type of advantage that betweenness highlights is the ability to broker between nodes that are otherwise separated: an advantage that can be sorted into information control and enhanced creativity as those nodes, on the one hand, control the information flowing between groups and, on the other, draw from pools of resources otherwise disconnected.

Yet, the structure of the co-authorship network should not be the only data needed to predict the impact of a paper that is also affected by its content and its position in the literature.

It comes useful the work by Kaplan and Vakili (2012) who introduced a new type of measure of creative performance on the basis of the content alongside the traditional citation count that reflects peers' recognition. They introduce topic generation as a measure of creativity. An article generates a new topic when its content is diverse from that of the articles already present in the literature. For instance, if in the literature there are many articles on issues related to draughts, and one article introduces problems linked with floods, this article will probably be topic generating if much of its lexicon is new. To determine the creation of a new topic, an automated textual classification is run through the whole set of documents in order to extract consistent topics and attribute a proportion of each of them to each document in the set(Blei, Ng, & Jordan, 2003). Similarly, we identify articles that brings new topic-combinations to the literature.

In this work, we delve into the problem of the recognition and distinction of reasons behind the cognitive and/or *traditional* success of an article, and we introduce network dimensions, namely co-authorship and bibliographic coupling (Newman, 2010), to enable us to control for a larger array of possible explanations. Specifically, we try to understand what is associated with the citation count, answering the following research questions: " (1) how the introduction of a new topic and the new combination of topics in a literature influence the citation counts? (2) How the network structure of co-authors influences the citation count? (3) And the knowledge sources of a paper influence the citation it receives?

To answer these questions, we use a dataset of 3343 publications from 1989 to 2010 on climate change vulnerability, see Table 4, that spans across several research fields and created multiple streams of literature such as Climate Change Adaptation and Disaster Risk Reduction.

Table 4 - Keywords

Vulnerability & risk assessment
Vulnerability & risk management
Vulnerability & adaptive management
Vulnerability & water resource management
Vulnerability & climate
Vulnerability & climate change
Vulnerability & climate change adaptation
Vulnerability & disaster risk
Vulnerability & disaster risk reduction

The article has the following structure: we first develop the hypothesis; second, we describe how to we extracted the thematic structure of the articles and the networks of co-authorships and references, and then we discuss the results. Lastly, we derive theoretical implications.

THE IMPACT OF SCIENTIFIC RESEARCH

Using citations to measure the impact of research has decades of history (Garfield, 1955; Garfield, 2006) and has produced substantial knowledge on the assessment of the most relevant journals in each scientific field.

In terms of citations, one of the most striking findings is that about 80% of citations is produced roughly by 20% of the articles (Garfield, 2006; Mazloumian et al., 2011). More specifically, very many articles have a very small number of citations (0 or 1) and tend to be forgotten in the ocean of publications, while an exceptionally small number of them become classic cornerstones of scientific research and get cited an incredibly high number of times (even beyond 10,000). Citations are used to determine the success of an author such as in the case of the h-index that sums the number of papers published by an author that received more than a given amount of citations (Hirsch, 2005). Citations are also sensitive to the disciplinary contexts, varying in terms of number and time to reach half of the overall citations (Wendl, 2007), and to the type of article: review articles are more cited (Ioannidis, 2006; Wendl, 2007).

Generative creativity

Notwithstanding the diffusion of the citation count as a proxy for the scientific impact of an article, how is it affected by the type of contribution? For instance, are novel contributions are rewarded?

It is not just an axiomatic claim that the first papers published in a field should be the most important and thereby they will be highly cited (Redner, 2005), evidence in a specific research stream showed that when it is possible to clearly establish the beginning of a field, the first articles significantly benefit from a higher citation rate (Newman, 2009). Therefore, when an article introduces a new topic into the scientific discussion, it should carry with itself the possibility of opening a new research stream, lead future research that will necessarily cite something they build upon, and thereby benefit from an advantage of citations over the followers. The degree of innovativeness of such articles should be reflected into their text, with respect to what is already present in the extant literature.

To our knowledge, such an advantage of the forerunner holds also among patents within the same technological area. Those which first introduce a new technology in the field get cited twice as much as the others (Kaplan & Vakili, 2012). Therefore, there is a reason to believe that when a new topic is first brought into the extant scientific literature, the article will receive a higher number of citations

Hypothesis 1: the introduction of a new topic into the extant scientific literature will positively affect the citation received by the article.

Combinatorial creativity

The power of association of distant ideas has been acknowledged long ago by Aristotle who stated that the 'hallmark of a genius' is to be able to recognize similar into the dissimilar and thereby to be able to associate distant things. This idea is stressed also by important thinkers of different kinds such as Coleridge, Einstein, and Poincaré who claimed that their insights stemmed from combinatory activities (Mednick, 1962). Creative insights in an individual mind are attributed to two mental operations: associative thinking and analytical thinking. The former is the process by which links between loosely associated ideas are identified, whereas the latter is the process by which a synthesis on a common structure is created between different – at times distant – domains (Fauconnier & Turner, 1998; Thagard & Stewart, 2011).

Novel combination of knowledge may lead to breakthrough discoveries, and tend to be credited more and for a longer period of time. Empirical evidence show that the power of associating ideas even of related domains fosters scientific creativity and knowledge advancement also in laboratory meetings (Dunbar, 1999) , and also connecting authors with different expertise may lead to the exploration of unconventional scientific routes tested to provide innovative solutions (Biscaro, Comacchio, & Warglien, first chapter).

A second type of creativity is of a combinatorial type (Frigotto & Riccaboni, 2011) and it is achieved when an article first introduces a combination of topics within the dataset. Both inside and outside the scientific domain, creating new pathways across distant knowledge bases is an effective way to generate innovative, out-of-the-box ideas (Burt, 2004; Frigotto & Riccaboni, 2011) that may translate into new paradigms and successful solutions. Thereby, we expect that bringing to the literature a new combination of topics will generate a positive impact on the citations received by the article.

Hypothesis 2: the introduction of a new topic combination into the extant scientific literature will positively affect the citation received by the article.

Where to stand: the co-authorship network structure

Part of the success of an article is determined by the fact that its ideas have the possibility to flow and be communicated to others, thereby the social relationships that the authors constructed with other authors over time represent a means by the information diffusion (Borgatti, 2005). Given that ideas are goods that can be shared simultaneously through emails, and other means of communications, to direct contacts, the number of people that can be reached almost effortlessly by co-authors is given by the number of connections they have established. Yet, it is not just the number of different connections that matters, because when authors have established many links but people that are all part of an isolated group, the diffusion will be limited to the number of co-authors in the group, because there exist no social tie that can facilitate the spread of the information beyond the boundaries of the group. Instead, when social ties between groups of co-authors are created, ideas can be effortlessly diffused beyond the co-authors boundaries. Thereby, when one co-author brokers between groups, the diffusion of an idea is not bounded by a limited number of social ties, but it can reach more co-authors. And the less numerous is the number of social relationships between an authors and all others in the field, the faster the idea can be diffused throughout the whole network.

Social relationships play a key role also in knowledge generation (Burt, 2005). A greater number of social connections enables co-authors to draw on a larger pool of knowledge, to receive multiple feedbacks from peers, and this allow co-authors to enhance the quality of their scientific production. Similarly the co-author who brokers between groups can be influenced by their different specializations, interpretations, and information (Burt, 2005). Dealing with different research communities may help brokers to move ideas, practices and methods that are mundane in a group to a context in which they are new. This combination may generate valuable knowledge (Hargadon & Sutton, 1997; Perry-Smith, 2006).

Thereby we expect to see a positive effect on the scientific impact of an article as the number of co-authors' social relationships increases. Similarly we expect a positive effect from a higher brokering position of the co-authors. Such advantages are both related to faster diffusion and enhanced creativity. Only related to the diffusion process, we expect that closer the coauthors are to all other coauthors in the field, the faster will be the diffusion of the idea, thus benefitting the impact of the scientific article.

Hypothesis 3a: the number of social relationships of the article's co-authors is positively associated with the impact of the scientific article..

Hypothesis 3b: the brokering position of the article's co-authors is positively associated to the impact with the scientific article.

Hypothesis 3c: the smaller the number of social connection between the article's co-authors and all other authors in the field, the larger will be the scientific impact of the article.

Where to build: bibliographic coupling network structure

Article' references signal information on the article position in the literature by highlighting the type of literature to which its co-authors want to contribute and by displaying from which literature they borrow knowledge. Connections among remote pieces of knowledge may lead to non-trivial solutions to problems that are otherwise difficult through conventional analytic thinking (Gick & Holyoak, 1980), and this have support both within and outside the scientific context (Dunbar, 1999; Sawyer, 2012). Yet, deep knowledge of a scientific context is a necessary condition to import a novel solution from outside, because the combinatorial thinking is generated through the mapping between elements of two different contexts (Fauconnier & Turner, 1998) and also because of the technical experience that is necessary to validate possible solutions (Mumford & Gustafson, 1988). For example, novel scientific knowledge may emerge when scientists bring in the field techniques and methods that are standard in another domain. Such as the example of Helmholtz who invented the ophthalmoscope to analyze the eye by combining his passion for optics into his training in physics.

Mapping the sources of knowledge of a scientific article gives the possibility to verify the combinatorial generation of knowledge. Knowledge combination can be studied through the analysis of references (Rafols & Meyer, 2010), specifically by establishing relations between articles when they share at least a reference, it is possible to generate a representation of a literature based on the knowledge sources. For example, a new theoretical framework introduced in an article, which unifies fragmented pieces of scientific literature prior to its publication, will generate a node in the network of the extant literature that brokers between the set of other articles that use different fragments of the scientific literature. Thereby, the article will appear between detached groups of nodes, thus creating a short path between them. It is expected that such article will benefit from blending ideas, methods or techniques from different strands of literature will be more likely to generate a novel and original contribution that will positively influence its scientific impact.

Hypothesis 4a: the more an article draws on fragmented pieces of knowledge, the larger will be its scientific impact

Furthermore, the relevance of a debate to which authors want to contribute is also a condition associated with the impact of the creative effort.Within a given literature, there are niches as well as mainstream topics, and they are such because of the different size of audience and contributors. Sharing parts of the references with other articles signals that co-authors want to contribute to and interact with a particular research community identified by those articles. The number of articles with which an article shares its references identifies the size of the audience to which the authors what to communicate. Although the importance of problems may change over time and a niche may turn into mainstream, we can conjecture that path dependence is what most likely occurs. Thereby the size of the strand of literature with which an article shares its references will be positively associated with the impact of its scientific contribution.

Hypothesis 4b: the size of the strand of literature with which an article share its references is positively associated with its scientific impact.

METHODS

Discovering topics from articles

Scanning and separating large collections of documents according to their underlying themes can be grueling and extremely time consuming. However recent developments in computer science applied to natural language processing (Blei, 2012; Blei et al., 2003; Chang & Blei, 2010) have reached the ability to discover the hidden thematic structure of documents through the analysis of their only observables: words and their allocation in a corpus of documents.

The basic idea that a document can reflect one or more topics. For example an article on vulnerability may have some parts of mathematics, some others of evolutionary biology, and some of economics. As a consequence a document can be seen as a distribution over topics, whereas a topic is a probability distribution over the whole set of words in the vocabulary (i.e., the set of words adopted in the whole collection of documents in the dataset). In natural language processing, the idea of describing the contribution of different topics to a document is commonly modeled bearing in mind that a word w_i is extracted randomly from a distribution over word given by the topic $P(w_i|z_i = j)$ multiplied by $P(z_i = j)$ that is probability of choosing a word in a definite topic *j* (Blei & Lafferty, 2006; Blei & Lafferty, 2007; Blei et al., 2003; Griffiths & Steyvers, 2004). And if there are *T* topics, the probability of the *i*th word is given by $P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j)$. The intuition is that $P(w|z)$ gives the idea of the importance of words in topic, whereas $P(z)$ is the probability to find a topic in a document.

The origin of the generative Latent Dirichlet Allocation (LDA) algorithm underpinning topic modeling is to trace back in genetics as it has been used to model allele frequencies and to allocate individuals to different population groups according to their expressions. In the analogy of topic modeling adapted to texts, individuals are like *documents* because they contain a combination of expressions or alleles, that are observable *words* and the population group is the latent variable that must be inferred, therefore they are the *topics* that are distributions over the alleles (Pritchard, Stephens, & Donnelly, 2000).

The LDA algorithm requires to specify the number of topics, then the algorithm tries to maximizing the joint probability $P(\mathbf{w}|\phi, \theta)$ that describes the equation $P(w_i) = \sum_{j=1}^{T} P(w_i|z_i = j)P(z_i = j)$ computed for all words in a document as well as for all documents. ϕ is a set *T* multinomial distributions of each word in the document and describes the probability of each word to be generated by the topic distribution. Whereas θ is a set of *D* multinomial distributions over the *T* topics (for a simple and more exhaustive explanation see Blei, 2012).

LDA assumes that a document is a “bag of words”, that is the order of words is irrelevant, an assumption that is unrealistic for language generation, but it is fine for unveiling the hidden structure of topics. Moreover, LDA assumes that the order of documents is also irrelevant and this assumption is ok if applied to a set of documents that are written in a relatively short period of time. When they span over many years or centuries, it is possible to define topics as series of distributions over words and see how they change over time (Blei & Lafferty, 2006).

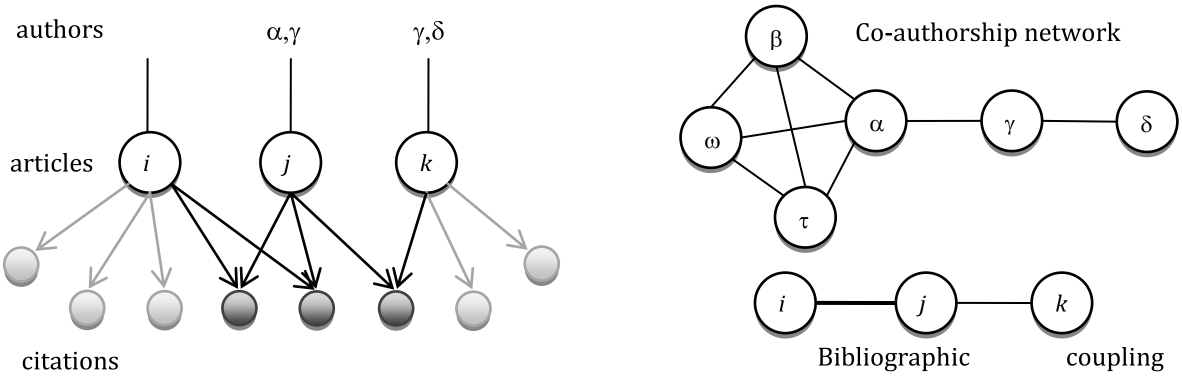
Among the applications, topic modeling has been adopted to draw relations among scientific articles based on their thematic similarity (Chang & Blei, 2010), and to categorize a dataset of technological patents in similar thematic group (Kaplan & Vakili, 2012). In this research, we adopt topic modeling to categorize the whole corpus of scientific articles and sift out the non-relevant ones avoiding to compare articles of different disciplines.

Networks: bibliographic coupling and collaboration networks

The structure of a scientific literature and that one of collaboration between authors can be analyzed through networks that are based on the mathematical mapping of relations (edges) between dyads of elements (vertices or nodes).

From a set of scientific articles, it is possible to establish relations among different attributes, e.g., references, co-authors, keywords. In this research, we focus on two types of networks: the network of collaborations between authors (co-authorship network), and a network of shared references (bibliographic coupling or BC network). In the BC network, the articles are the vertices and an egde is established when they have at least one shared reference. The weight of the edge is given by the number of shared references. Analogously in the co-authorship network, vertices are the authors and the edges are established between vertices who co-authored an article. The weight of the edge is given by the number of articles written by a dyad of vertices.

Figure 9 – Example of bibliographic coupling network and and co-authorship network stemming from a set of three articles.



In Figure 1, it is represented how to construct the two networks starting from a set of three articles *i*, *j*, and *k*. Articles *i,j* share two references, while *j* and *k* share one, thus in the BC network, node *j* is linked with and *i* with an edge whose weight is equal to two, and with *k* with an edge of weight one. Node *j* stands in between nodes *i* and *k*, as it shares some of the literature with *i*, and some with *k*, while the former two articles do not share the any references. In the co-authorship network, through the article *i*, the author α establishes an edge with β, ω and τ, and through *j* with author γ.

While co-authorship networks are widely used to study scholarly collaborations, bibliographic coupling networks are a better alternative to co-citation networks that establishes relations between the all articles in the reference list, because co-citation leaves out most of the articles that are not highly cited, and generates a larger network because, especially when the literature upon which is fragmented (in our case we would have over 250,000 nodes, instead of 3343) and makes it computationally easier to analyze. Moreover, bibliographic coupling can be computed as soon as the paper are published and do not vary over time, unlike cocitation (Newman, 2010).

One of the disadvantages of the bibliographic coupling is that the strength and number of relations depends on the size of the reference list, giving a positive bias towards those articles who have a larger reference list, as they are more likely to share any of their reference with other articles. However, this can be controlled by normalizing for each paper the edges' weight over the size of the paper's reference list, operation that makes the graph directed.

Research Setting

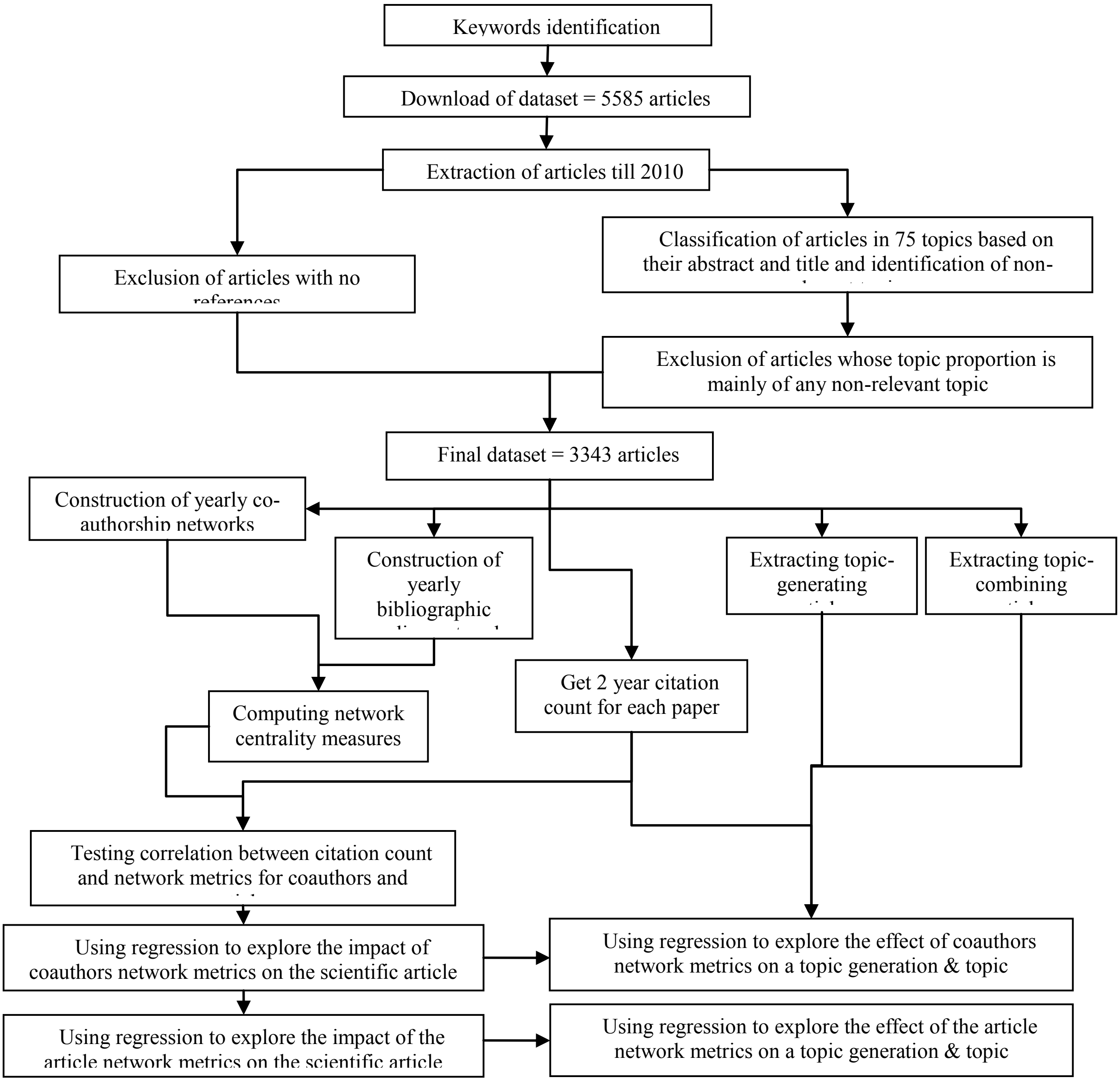
Our hypothesis were tested in a dataset of 3343 scientific articles retrieved by search in the Web of Knowledge performed on April 03 2013 database with keywords on Vulnerability, in Table 1.. The context of vulnerability to climatic events and climate change has been chosen, because of its great variability and of its importance in the climate change literature. It generated contributions focused on a multitude of aspects that vary from strategies to adapt and anticipate possible consequences of a natural event to strategies to better cope with their effects in a multitude of different ecosystems, affected by diverse types of natural events (floods, draughts, storms, heat waves, and extreme wind, among the others), and mediated by various morphological, geological, and social conditions. Moreover, articles in the literature vary in the unit - and level - of analysis, as well as in the methodology.

The final dataset of 3343 articles comes from a larger dataset of 5585 papers published between 1985 and 2013, whereby we discarded those published after the 2010 for the measure of scientific impact chosen (2 complete years of citations), see Figure 2. We classify articles based on the latent structure of topics[6] extracted through the textual information retrieved in the search: abstracts and titles. The algorithm requires to arbitrarily set the number of topics, that was set to 75, a large number who took into account the variability of the themes discussed in the field: types of environments, receptors, and unit of analysis. Topics were coded by an expert in the field (with more than three publications) who labeled each topic on the basis of its 20 most likely words. Such coding was then validated by two other experts, who also identified nine non-relevant topics that we used to exclude the articles whose majority of the content was in one of these (topic proportion > .50; 271 articles excluded). Moreover, for the impossibility to compute the position in the BC network, we excluded the papers with no references and also those without a sufficient number of

[6] We used the Stanford Topic Modeling suite.

words in the title and the abstract (less than 5), for the impossibility to find their thematic structure,. This process brought the dataset to a population of 3343 articles published between 1989 and 2010. We then created the BC and co-authorship weighted networks on temporal slices that kept fixed the first observation in the dataset and moved forward by one year (e.g., 1989-2002, 1989-2003, 1989-2004) and for each network we extracted network measures by means of the library ‘igraph’ in R (Csardi & Nepusz, 2006). The temporal slices enabled us to observe the conditions whereby articles are published both in terms of co-authorship collaborations and in terms of position of the article in the literature, and to observe the dynamics of the networks. Lastly, we performed statistical analysis by means of negative binomial regressions.

Figure 10 - Mapping the process of data analysis



Measures

Scientific impact. The only dependent variable is the articles' scientific impact and it is measured by means of the number of citation received by an article in a fixed time window of two years after the year of publication (hereafter *citation count*). Such a choice has been made not to favor old articles that can be cited for a longer period of time and have a larger number of potential citing articles. We decide to exclude the citations received during the year of publication, as that would create a bias in favor of the articles published early in the year. The fixed citation time window is shown to be a good measure to determine the scientific impact within a discipline, whereas it is flown for cross-disciplinary comparisons, because of different maturity times (Dorta-Gonzalez & Dorta-Gonzalez, 2013), for this reason we used the topic categorization to exclude non-consistent articles.

Topic generation. An article is generates a topic when the topic proportion is for the first time larger than 0.25 (up to a maximum of 1). When in the same year and in the same topic, the threshold is surpassed by multiple articles, all of them are classified as topic-generating. A similar operationalization has been adopted by Kaplan and Vakili (2012) who considered a threshold of 0.20. Topic generation is a dummy variable that assumes value 1 when an article generates a new topic, and 0 otherwise.

New topic combination. Similarly to topic generation, articles are topic combining when they establish a combination of topics that was not present in the dataset until the year of publication. We adopted the same rule used to attribute topic generation: a topic proportion beyond the threshold of 0.25 determines the presence of the topic in the article and the rule of the multiple attribution of a topic combination is given by all those articles that introduce the same topic-combination in the same year.

For robustness, results do not significantly change when checked with thresholds of .20 and .30.

Degree centrality is the count of the first neighbors of a node. In a co-authorship network it equates with the number of co-authors with whom any author has collaborated at the year *t* (hereafter *author degree)*. For the BC network, we use the *degree proportion* that is the degree centrality divided by the number of nodes in the network at year *t*. Thereby it is the proportion of articles in the dataset with which an article shares at least one reference.

Closeness centrality considers how close the node is to any other node in the network. Therefore a high closeness score indexes a short distance between the node on which it is computed and any other node. For a single node *i*, it is the inverse of the mean distance of the geodesic (shortest) path *g* to any other.

$$\text{Closeness Centrality } (\text{Au}_i) = \left(\frac{1}{n-1} \sum_{j(\neq i)} g_{ij} \right)^{-1}$$

Betweenness centrality of a node *i* is computed summing the number n_{st}^{i} of geodesic paths between any two nodes *s* and *t* that pass through *i* over the total number g_{st} of geodesic path between the two nodes. This gets normalized by dividing by the number of ties between any other two nodes.

$$\text{Betweenness Centrality } (\text{Au}_i) = \left(\frac{\frac{\sum_{st} n_{st}^{i}}{g_{st}}}{(n-1)(n-2)} \right)^{-1}$$

An author *i* with a high betweenness centrality score is in the shortest path between many other actors, and he or she will benefit from brokering advantages, especially when these other actors are disconnected if vertex *i* is removed.

As our unit of analysis is the article, we are interested in network metrics of co-authors that can be mapped into the article. Thus we retrieve the maximum value for each measure as the effects of the most influential coauthor who transfers the highest value of authority to the paper (Mazloumian et al., 2011).

All metrics are computed on yearly slices to capture the values at the year of publication.

Control variables

Size of the citing literature. The measure of performance is given by the number of citations received by an article in the following two complete years from the publication date. They depend on the size of the universe of article from which citations are drawn, therefore there is a need to control for the size of this expanding universe. As a proxy of the expanding universe of articles, we take the number of articles in the dataset two years after the year of publication of each article. Although we recognize that this universe must not be exact universe of citing articles, it provides the sense of growing attention towards the topic of vulnerability and, secondly, it is a monotoniccally growing body of scientific literature, such as the entire universe of scientific literature. For robustness, we also tried two different versions of the metric: (1) inflated the measure with a relatively large fixed number representing the articles outside the dataset (10,000) that could be interested in referring to articles within; (2) we increased that fixed number by a 4.1% each year as it is the rate of growth of scientific articles in the period between 1990-2007 (Larsen & von Ins, 2010). Results proved to be robust.

Number of authors. The number of co-authors is positively correlated with the citation received, see Table XX, and there may be multifold reasons: a paper with multiple coauthors is more likely to be more complex in terms of knowledge sources, as it required the work of multiple actors; the quality of the content can also be enhanced by the *labor limae* that can be performed by multiple hands; furthermore, the dissemination of the ideas included in the article can go spread in the coauthorship network starting from multiple starting nodes. Thus, we sift out the effect by including the discrete number of co-authors in the regression.

Experience in the field. Experience is associated with a higher level of specialization, knowledge of the relevant problems in the field and with a deeper ability in publishing and diffusing ideas (Maske, Durden, & Gaynor, 2003). Thereby expert authors have both cognitive and reputation advantages as they are more known within the field, have a better knowledge on the conference to atten and have attended a greater number than their less-expert peers, and this should translate into better quality knowledge production which is also disseminated better. We operationalize the experience in the field for any article *i* by measuring the number of articles in the dataset published by each co-author prior to the publication of *i*. Among these number, we chose the maximum.

Citing review bias. In several scientific disciplines there is a high concentration of reviews among the most cited articles (Ioannidis, 2006). The review has the purpose of setting a line among scattered articles in the field and per se, this is already a good contribution to science. However, sometimes it is difficult to discriminate between review and not review based solely on the number of references, because to introduce a new conceptual framework to analyze new data, often authors draw on multiple contributions in the extant literature. Yet, to control for the citing review bias, we clustered articles in two groups (review, non review) based on the number of article cited in their reference list. The two groups have significantly different means ($\mu_{\text{review}} > \mu_{\text{non review}}$, $p < 0.001$), thus we created a dummy variable with value 1 if the article is a review, and 0 otherwise.

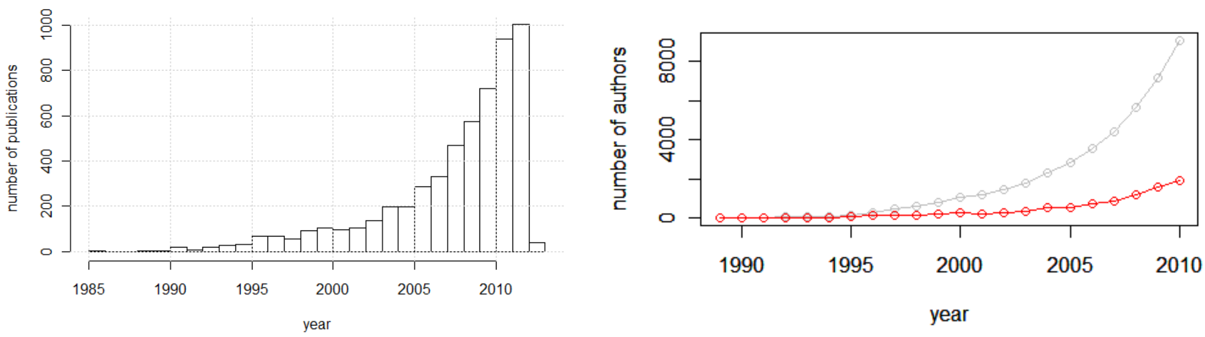
RESULTS

In this section, we first provide an overview of the data, then we answer the three research question and show how the *citation count* is affected by the topic generation and new topic combinations, by the structure and positions of co-authors within the co-authorship network, and by the position of the article in the literature.

Summary statistics

After a first publication 1985, then excluded because without references, the literature starts building in the early 90s, and has a steady increase over the years, as displayed in Figure 3 (left). From 2004, the number of papers increases in a steep-log phase. Analogous is the trend of co-authors displayed in Figure 3 (right). In grey the cumulative number of authors in the dataset, while in red the number of new authors entering in the dataset at any year. Most articles are written by 3 authors (mean 3.404, $\sigma = 3.021$, min = 1, max = 57), have on average 48.65 references ($\sigma = 34.537$, min = 1, max = 398), and receive on average 4.95 citations in the first two years after the publication ($\sigma = 9.890$, min = 0, max = 273), however the distribution of citation is skewed on the right.

Figure 11 – (left) Number of articles retrieved per year on "vulnerability" in the ISI Web of Knowledge. (right) Number of different authors: in red the number of different authors publishing each year, in grey the cumulative sum of different authors contributing to the field.



The correlation coefficients between the variables presented in Table 2 use Spearman correlation because of the non-normal distribution of most variables in the study. The correlation coefficients between the citation counts and the centrality measures computed on the co-authorship nodes (*Author degree*, *Author betweenness*, and *Author closeness*) show that author degree have a stronger positive association with the citation counts than *author betweenness*, whereas the closeness metric is slightly and negatively associated with the citations received. Degree and betweenness centralities are strongly and positively correlated (0.75 with p < 0.001), as in most networks (Goh, Oh, Kahng, & Kim, 2003).

Table 2 - Spearman Correlation Table with significance level

	Citations	(2)	(3)	(4)	(5)	(6)	(7)	(8)
(2) Author degree	0.33†							
(3) Author betweenness	0.29†	0.75†						
(4) Author closeness	-0.13†	-0.14†	-0.08†					
(5) Article degree proportion	0.22†	0.19†	0.30†	0.08†				
(6) Article betweenness	0.30†	0.24†	0.29†	-0.13†	0.70†			
(7) Number of authors	0.22†	0.74†	0.33†	-0.10†	-0.02	0.08†		
(8) Size of Literature	0.18†	0.26†	0.22†	-0.97†	-0.02	0.17†	0.15†	
(9) Author's experience	0.27†	0.71†	0.88†	-0.12†	0.33†	0.33†	0.20†	0.25†

Levels of significance: † p< .001.All measures are computed at the year of publication of each paper.

We also find that *articles degree proportion* and *article betweenness* are positively correlated with the citations received (.22 and .30 respectively with $p < .001$) and between each other (.70 with $p < .001$). Also control variables such as the *number of authors*, the *size of literature* and the *author's experience* are positively and significantly correlated with the citation count.

In the next paragraph we present the results of the regressions computed on three models. In model 1, we replicate what other researchers have done in the past (Uddin, Hossain, & Rasmussen, 2013) to study the structural effect of the co-authorship network on the *citation count*. In model 2, we add the dummy variables of *topic generation*, and *new topic combination*. Model 3 also includes the BC network measures. To control for outliers, we exclude the first three observation from the regressions.

Author closeness takes abnormally large values due to the initially small size of the network (the first three observations take values of 1 and .33, whereas the fourth largest observation of .038).

$$\text{citation count} = \beta_0 + \beta_1 \text{au. degree} + \beta_2 \text{au. betweenness} + \beta_3 \text{au. closeness}$$

$$+\beta_4 \text{new topic generation} + \beta_5 \text{new topic combination}$$

$$+ \beta_6 \text{art. degree proportion} + \beta_7 \text{art. betweenness}$$

$$+\beta_8 \text{number of authors} + \beta_9 \text{size of literature} + \beta_{10} \text{review} + \beta_{11} \text{au.'s experience} + \varepsilon$$

The impact of topic generation and topic combination.

All regressions are performed with negative binomial models that is proper for count data models and has no specific assumption, unlike the Poisson and the zero-inflated Poisson. Poisson assumes that the mean and the variance of the dependent variable should be equal, while in our dataset they differ significantly ($\mu = 4.952$, $\sigma^2 = 97.8116$). As for many articles in the dataset *citation count* take the value 0, we still prefer the negative binomial model to a zero-inflated Poisson, because the latter model assumes that many of the observations that take value zero are drawn from a different distribution, and thereby they will never be cited. In our case, there is no theoretical reason to assume that non cited articles come from a different distribution.

Table 3 – Results of Regression Analysis

	model 1		model 2		model 3	
	Estimate		Estimate		Estimate	
Intercept	0.738	†	0.734	†	0.701	†
	0.055		*0.055*		*0.056*	
Author degree	0.021	†	0.021	†	0.021	†
	0.003		*0.003*		*0.003*	
Author betweenness	-61.6	**	-60.61	**	-60	**
	19.6		*19.6*		*19.59*	
Author closenesss	-18.510	.	-12.190		-11.13	
	9.516		*10.750*		*10.85*	
Topic generation			-0.198		-0.191	
			0.159		*0.159*	
New topic combination			0.063		0.065	
			0.072		*0.072*	
Article degree proportion					-0.002	
					0.74	
Article betweenness					0.682	**
					0.249	
Number of authors	0.079	†	0.080	†	0.082	†
	0.007		*0.008*		*0.008*	
Size of Literature	4.41E-05	†	4.32E-05	**	4.81E-05	†
	1.33E-05		*1.34E-05*		*1.35E-05*	
Review	0.726	†	0.724	†	0.674	†
	0.053		*0.053*		*0.057*	
Author's Experience	0.025	.	0.025	.	0.019	
	0.013		*0.013*		*0.014*	

Signif. Codes: '†' p < 0.001; '**' p < 0.01 ; '*' p < 0.05; '.' p < 0.1. Standard errors in italics Number of observations: 3340

Model 2 in the regression table, Table 3, shows that generating a new topic (*topic generation*) within the literature does not benefit in terms of citations count (-.198, $\sigma = .16$, p-value = .21). For the large p-value, there is no possibility to claim that *topic generation* is detrimental. Most of the topic generating articles are concentrated in the early years of the vulnerability literature, thereby it may be expected that the overall number of citations articles in the first two years will be lower than for articles cited in later years, nevertheless, despite controlling for the size of literature in different ways, results remain negatively associated. When we regress the *citation count* also on the bibliographic coupling centrality measures, *topic generating* remain negatively, but not significantly, associated. Thereby we cannot reject the null hypothesis on topic generation, as there is no significant evidence that articles that introduce new topics in the literature are cited differently from others. A similar conclusion can be derived for hypothesis 2 that cannot be supported: *new topic combination* does not provide significant change in the *citation count*. In sum, data do not support hypotheses 1 and 2.

The impact of co-authors network position on the article's citation.

Author degree has a slightly positive but highly significant coefficient (0.021 with p-value < $1.88 \cdot 10^{-13}$) and such positive association on citations remains when we add also the information on topic generation (model 2) and on the BC network (in model 3). This result shows that articles written by authors who have established more social-relationships (co-authorships with other authors in the dataset) tend to be cited more. Surprisingly, and different to other cases (Uddin et al., 2013; Yan & Ding, 2009), *author betweenness* is negatively and significantly associated with the citations received by the paper. The coefficient is large, and thereby of a non-intuitive interpretation, because normalized betweenness scores take small values (min 0, max 0.025, but the distribution is right skewed and 75% of the values are below 6.06E-07). This results shows that bridging between groups of co-authors is not positively associated to *citation count*.

Author closeness's coefficient is negatively associated to the *citation count*, but not significantly. This is due to the lack of collaboration between groups until the most recent years and we can see it by the fact that most authors have just relationships of degree 1and are in isolated groups of co-authors. When we compute the same regression of model 3 in the presence of a tangible giant component[7] (from 2008, see Table 4), only with the 1683 articles published between 2008 and 2010, *Author closeness*'s coefficient becomes positive and significant (0.004 with p-value = 0.008). This results says that closeness is positively associated with citation count, and this could be due to the fact that the direct form of social interaction – the co-authorship – facilitates the diffusion of ideas.

Curiously, author betweenness remains negative and significant also when the giant component is large (after 2008).

Table 4 - Number of Authors in the Giant Component (GC) with respect to the total

	1998	2000	2002	2004	2006	2008	2010
Authors in the field	594	1003	1429	2222	3461	5398	8590
Authors in the GC	79	94	98	231	723	1565	3589
Percentage in the GC	13.30%	9.37%	6.86%	10.40%	20.89%	28.99%	41.78%

In summary, data support hypothesis 3a showing that articles written by co-authors with more social connections tend to be cited more. Hypothesis 3b is not supported, in fact co-authors who brokers among groups of co-authors have a significantly negative impact on citations. Hypothesis 3c - in model 3 – is partially supported.

The impact of the article network position on its citation count.

Regressions, in model 3,show that *article betweenness* is positively associated to the citation count (0.2 p-value < 0.01), thereby showing that articles that build upon fragmented, not well connected, strands of literature tend to be cited more than others. This highlights the fact that looking for relationships among theoretical arguments and finding connections to remote data is rewarded in terms of citations.

[7] The largest component of vertices that are connected to each other.

Article degree proportion instead does not impact significantly the *citation count* that is confirmed to the coefficient that is close to 0. Therefore, in this dataset, sharing references with a large number of other articles in the extant literature does not increase the number of citations received, suggesting that there is no apparent relation between recognizing an argument already well spread in the literature or acknowledging very popular prior research and the impact of the scientific production.

In summary, hypothesis 4a is supported, while hypothesis 4b is not supported.

Among the control variables, the number of co-authors, the type of article (review or non-review) and the size of literature are all positively and significantly associated with more citations, and especially review articles have a large positive effect on citations. When we control for the position of the article in the literature, the author's experience loses significance, showing that, at least in this literature, the number of articles published in the field is not a good predictor for the success of the next article.

DISCUSSION AND CONCLUSION

The inspiration for this research came from John of Salisbury's analogy. As dwarves stand on giants' shoulders in order to look further and do produce good research, in this paper we show that it standing on good shoulders matters. Or in other words, deciding with whom to collaborate may improve the citations of the article. Not only that, we show that the efforts of scientists who draw establish new connections between scientific pieces of literature tends to be more cited by future researchers.

To perform this analysis, we combined methods of network analysis with topic-modeling (Blei et al., 2003), a machine learning technique to process natural language, in order to extend – and partly provide opposite evidence to – the recent bibliometric results who highlighted the positive effect of some structural position of co-authors in the co-authorship network on the citation received by a paper (Uddin et al., 2013; Yan & Ding, 2009). Specifically, we address three research questions: " (1) how the introduction of a new topic and the new combination of topics in a literature influence the citation counts? (2) How the network structure of co-authors influences the citation count? (3) Do the knowledge sources of a paper influence the citation it receives? To answer these questions , we retrieved the articles published in the scientific literature on climate change vulnerability that was used to construct dynamic networks of co-authors, of knowledge sources, and to identify latent topics. We first show that *topic generation* as well as *new topic combinations* do not significantly affect neither positively, nor negatively the citations received in the first two years after the publication. For *topic combination*, the vast nature of the dataset allowed the identification of distant and probably incongruent topics such as draughts, river basins and arctic, whose combination in the same article may signal authors' confusion rather than a provision of a smart contribution. Notwithstanding cause-effect relations might be possibly drawn between distant topics, peer recognition comes from deep analytical analysis and counterintuitive results, and these are not captured by the topic modeling algorithm. However, the fact that topic generating articles are concentrated in the early years of the literature coupled with the sparse co-authorship network until the latest years are signs that the body of literature on vulnerability does not cohesively grow upon seminal contributions in which authors speak to each other. Instead, the articles in the dataset appear as belonging to different bodies of literatures (probably pertinent to the topics as pre-existing different literatures) which introduce the concept of vulnerability in different years. Both in terms of collaborations among authors and in terms of finding ways to unify fragmented strands of literatures, the emergence of the literature of vulnerability can be seen as a bottom-up process whereby scattered contributions find a cohesion over time.

In terms of the position of scientists in the co-authorship network, we show that *degree centrality* has a positive effect on the citation count, whereas different to prior results (Uddin et al., 2013; Yan

& Ding, 2009), we found that the *author betweenness* is negatively associated to citation counts. The result of the *degree centrality*, therefore the number of co-authors computed on the authors with most co-authors, can be interpreted both in terms of knowledge combination and in terms of diffusion of the ideas. In the former terms, higher degree scores entails that the authors have a larger knowledge endowment to eventually tap or simply have constructed a deeper and more profound knowledge ground due to their past and present relations, and they are able to exploit their social network to write more impactful articles. In terms of knowledge diffusion, the social network may ease the diffusion of ideas (Valente, 1996), thereby a larger immediate social network enhances the odds for the article to be cited in the next two years (including self-citations). Also the *author closeness*, when the giant component comprises a big portion of the authors in the field, is positively associated with the *citation count*. This may be interpreted as the possibility for the idea to spread through more rapidly and also because they, belonging to a larger component, are known by a larger number of people. Yet, by the nature of the literature on vulnerability that is itself fragmented, this finding need to be tested elsewhere. Doubts remain on the interpretation of the negative impact created by an increase of the *author betweenness*. If a two year time-window in which citations are counted were not enough to spread the idea beyond the co-authors, there would be no reason for *authors closeness* to have a positive impact on citations. More work has to be done to understand why there is no advantage to broker and the current data do not inform possible theoretical support for its interpretation. Thereby further studies may try to sieve if a negative effect of *author betweenness* is contingent to the dataset or if there is some theoretical reason underpinning it.

We find that knowledge sources have an effect on citations. Specifically articles that find ways to tie together fragmented pieces of literature benefit from more citations. While neither a specific benefit, nor a disadvantage in the citation count is given by mentioning in the reference list articles that are already cited by many others in the extant field.

Indeed, this work comes with its limitations. Our decision to attribute to each article the value given by the co-author with highest centrality prevented us from analyzing the information given by the heterogeneity of the set of co-authors. They may have different positions in the co-authorship network, therefore they may explore different knowledge sources and be themselves multiple – and distant – sources for the dissemination of the article. Moreover, considering only the network positions of co-authors prevented us from understanding how the nature of the collaboration, e.g., instance the composition of the team, the frequency of dyadic co-authorships, the entrance of new individuals in the team, and the experience of the new entrant are likely to affect knowledge combination within articles. For instance, how group compositions affects scientific productions? How researcher turnover affects knowledge production?

REFERENCES

Blei, D. M. 2012. Probabilistic Topic Models. ***Communications of the Acm***, 55(4): 77-84.

Blei, D. M., & Lafferty, J. D. 2006. ***Dynamic topic models***. Paper presented at the Proceedings of the 23rd international conference on Machine learning.

Blei, D. M., & Lafferty, J. D. 2007. A correlated topic model of science. ***Annals of Applied Statistics***, 1(1): 17-35.

Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. Latent Dirichlet allocation. ***Journal of Machine Learning Research***, 3(4-5): 993-1022.

Borgatti, S. P. 2005. Centrality and network flow. ***Social networks***, 27(1): 55-71.

Burt, R. S. 2004. Structural holes and good ideas. ***American Journal of Sociology***, 110(2): 349-399.

Burt, R. S. 2005. ***Brokerage and Closure: An Introduction to Social Capital: An Introduction to Social Capital***: Oxford University Press.

Chang, J., & Blei, D. M. 2010. Hierarchical relational models for document networks. ***The Annals of Applied Statistics***, 4(1): 124-150.

Chen, C. M. 2004. Searching for intellectual turning points: Progressive knowledge domain visualization. ***Proceedings of the National Academy of Sciences of the United States of America***, 101: 5303-5310.

Csardi, G., & Nepusz, T. 2006. The igraph software package for complex network research. ***InterJournal, Complex Systems***, 1695(5).

Dorta-Gonzalez, P., & Dorta-Gonzalez, M. I. 2013. Impact maturity times and citation time windows: The 2-year maximum journal impact factor. ***Journal of Informetrics***, 7(3): 593-602.

Dunbar, K. 1999. How scientists build models in vivo science as a window on the scientific mind, ***Model-based reasoning in scientific discovery***: 85-99: Springer.

Fauconnier, G., & Turner, M. 1998. Conceptual integration networks. ***Cognitive science***, 22(2): 133-187.

Freeman, L. C. 1979. Centrality in social networks conceptual clarification. ***Social networks***, 1(3): 215-239.

Frigotto, M. L., & Riccaboni, M. 2011. A few special cases: scientific creativity and network dynamics in the field of rare diseases. ***Scientometrics***, 89(1): 397-420.

Garfield, E. 1955. Citation indexes for science; a new dimension in documentation through association of ideas. ***Science (New York, N.Y.)***, 122(3159): 108-111.

Garfield, E. 2006. The history and meaning of the journal impact factor. ***JAMA: the journal of the American Medical Association***, 295(1): 90-93.

Gick, M. L., & Holyoak, K. J. 1980. Analogical problem solving. ***Cognitive Psychology***, 12(3): 306-355.

Goh, K. I., Oh, E., Kahng, B., & Kim, D. 2003. Betweenness centrality correlation in social networks. ***Physical Review E***, 67(1): 017101.

Griffiths, T. L., & Steyvers, M. 2004. Finding scientific topics. ***Proceedings of the National Academy of Sciences of the United States of America***, 101(Suppl 1): 5228-5235.

Hargadon, A., & Sutton, R. I. 1997. Technology brokering and innovation in a product development firm. ***Administrative Science Quarterly***: 716-749.

Hirsch, J. E. 2005. An index to quantify an individual's scientific research output. ***Proceedings of the National Academy of Sciences of the United States of America***, 102(46): 16569.

Ioannidis, J. P. 2006. Concentration of the most-cited papers in the scientific literature: analysis of journal ecosystems. ***Plos One***, 1(1): e5.

Janssen, M. A. 2007. An update on the scholarly networks on resilience, vulnerability, and adaptation within the human dimensions of global environmental change. ***Ecology and Society***, 12(2).

Janssen, M. A., Schoon, M. L., Ke, W. M., & Borner, K. 2006. Scholarly networks on resilience, vulnerability and adaptation within the human dimensions of global environmental change. ***Global Environmental Change-Human and Policy Dimensions***, 16(3): 240-252.

Kaplan, S., & Vakili, K. 2012. Identifying Breakthroughs: Cognitive vs. Economic.

Larsen, P. O., & von Ins, M. 2010. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. ***Scientometrics***, 84(3): 575-603.

Maske, K. L., Durden, G. C., & Gaynor, P. E. 2003. Determinants of scholarly productivity among male and female economists. ***Economic Inquiry***, 41(4): 555-564.

Mazloumian, A., Eom, Y. H., Helbing, D., Lozano, S., & Fortunato, S. 2011. How Citation Boosts Promote Scientific Paradigm Shifts and Nobel Prizes. ***Plos One***, 6(5).

Mednick, S. A. 1962. The associative basis of the creative process. ***Psychological review***, 69(3): 220-232.

Mumford, M. D., & Gustafson, S. B. 1988. Creativity syndrome: Integration, application, and innovation. ***Psychological bulletin***, 103(1): 27.

Newman, M. 2010. ***Networks: An Introduction***: OUP Oxford.

Newman, M. E. J. 2001. The structure of scientific collaboration networks. ***Proceedings of the National Academy of Sciences of the United States of America***, 98(2): 404-409.

Newman, M. E. J. 2009. The first-mover advantage in scientific publication. ***Epl***, 86(6).

Perry-Smith, J. E. 2006. Social yet creative: The role of social relationships in facilitating individual creativity. ***Academy of Management Journal***, 49(1): 85-101.

Polanyi, M. 2000. The republic of science: its political and economic theory. ***Minerva***, 38(1): 1-21.

Pritchard, J. K., Stephens, M., & Donnelly, P. 2000. Inference of population structure using multilocus genotype data. ***Genetics***, 155(2): 945-959.

Rafols, I., & Meyer, M. 2010. Diversity and network coherence as indicators of interdisciplinarity: case studies in bionanoscience. ***Scientometrics***, 82(2): 263-287.

Redner, S. 1998. How popular is your paper? An empirical study of the citation distribution. ***European Physical Journal B***, 4(2): 131-134.

Redner, S. 2005. Citation Statistics from 110 Years of Physical Review. ***Physics Today***, 58(6): 49-54.

Sawyer, R. K. 2012. ***Explaining creativity: The science of human innovation***: OUP USA.

Sun, X. L., Kaur, J., Milojevic, S., Flammini, A., & Menczer, F. 2013. Social Dynamics of Science. ***Scientific Reports***, 3.

Thagard, P., & Stewart, T. C. 2011. The AHA! experience: Creativity through emergent binding in neural networks. ***Cognitive science***, 35(1): 1-33.

Uddin, S., Hossain, L., & Rasmussen, K. 2013. Network Effects on Scientific Collaborations. ***Plos One***, 8(2).

Valente, T. W. 1996. Social network thresholds in the diffusion of innovations. ***Social networks***, 18(1): 69-89.

Wendl, M. C. 2007. H-index: however ranked, citations need context. ***Nature***, 449(7161): 403-403.

Yan, E. J., & Ding, Y. 2009. Applying Centrality Measures to Impact Analysis: A Coauthorship Network Analysis. ***Journal of the American Society for Information Science and Technology***, 60(10): 2107-2118.

Appendix 1. Programming Code C# - the Document Term Matrix

In Chapter 2, to create a document term (DT) matrix from a collection of text files (with a .txt extension) that are all located in the same folder, I used the following code:

```
// Main Routine
{
       //get all txt file in a folder
       string[] filelist = Directory.GetFiles(@"C:\folderName\", "*.txt");

       Dictionary<string, Dictionary<string, int>> countPerFile;
       //it maps from string to dictionary

       countPerFile = new Dictionary<string, Dictionary<string, int>>();


       // (1) Display all files.
       foreach (string filename in filelist)
       {
              Console.WriteLine("processing file " + filename);
              Dictionary<string, int> count; // it maps from string to integer
              count = new Dictionary<string, int>();
              countPerFile[filename] = count;
              scanFile(filename, count);
       }

       // (2) Function to create a complete dictionary

       Dictionary<string, int> globalEntries; // the list that contains all words
       globalEntries = new Dictionary<string, int>();

       int wc = 0; // word count

       foreach (string filename in filelist) // same as (1)
       {
              Console.WriteLine("processing files dictionary entry : " + filename);

              Dictionary<string, int> count; // it maps from string to integer
              count = countPerFile[filename];

              foreach (string wrd in count.Keys)
              {
                     if (!globalEntries.ContainsKey(wrd))
                     {
                            globalEntries[wrd] = wc++;
                     }
              }
       }

       showDictionary(globalEntries);

       // we create the DT matrix (DTM)

       int rows = filelist.GetLength(0); // define number of rows and columns
       int cols = globalEntries.Keys.Count;
       string[] entries = new string[cols];  // defining the length of the heading
                                             // this row will be filled with the
                                             // different words in the dictionary
```

```
        globalEntries.Keys.CopyTo(entries, 0);

        int[,] DTM = new int[rows, cols];

        // (3) filling the matrix
        // we read the file in the list in order from the first to the last
        for(int fileIndex = 0; fileIndex < filelist.GetLength(0); fileIndex++)
        {
                string filename = filelist[fileIndex];

                Dictionary<string, int> count;
                count = countPerFile[filename];
                foreach (string wrd in count.Keys)
                {
                        //  we find the corresponding index in globalEntries
                        int wordIndex = globalEntries[wrd];
                        DTM[fileIndex, wordIndex] = count[wrd];
                }
        }

        // (4) Create the output file
        string fileOut = @"c:\destinationFolder\outputFile.csv";
        StreamWriter sw = new StreamWriter(fileOut);

        // (5) use header to write the first row
        string header = "";
        for (int i = 0; i < entries.Length; i++)
        {
                header += entries[i];
                if (i < entries.Length - 1)
                {
                        header += ';';
                }
        }
        sw.WriteLine(header);

        // Reads the rows one by one from the SqlDataReader
        // transfers them to a string with the given separator character and
        // writes it to the file.

        for(int r = 0; r<rows; r++)
        {
                string strRow = "";
                for(int c=0; c<cols; c++)
                {

                        strRow += DTM[r, c].ToString();
                        if (c < cols - 1)
                        {
                                strRow += ';';
                        }
                }
                sw.WriteLine(strRow);
        }

        // Closes the text stream and the database connection.
        sw.Close();
}

// (Function showDictionary)
private void showDictionary(Dictionary<string, int> dizionario)
```

```
{
       string outString = "";
       foreach (string chiave in dizionario.Keys)
       {
              int valore = dizionario[chiave];
              outString += chiave + " ---> " + valore.ToString() + Environment.NewLine;
       }
       textBox1.Text = outString; // textBox1 <- preset display in C#
}

// (Function scanFile that pre-process text: stemming, eliminates numbers, etc.)

private void scanFile(string filename, Dictionary<string,int> count)
{
       StreamReader streamReader = new StreamReader(filename);
       string testo = streamReader.ReadToEnd();
       streamReader.Close();

       // preprocessing
       testo = testo.ToLower();
       testo = testo.Replace('1', ' '); // and other controls

       // eliminating stopwords
       string[] stopwords = File.ReadAllLines(@"c:\folderStopWords\stopWordFile.txt");
       string regexCode = @"(?<=(\A|\s|\.|,|!|\?))(" +
       string.Join("|", stopwords) +
       @")(?=(\s|\z|\.|,|!|\?))";
       Regex regex1 = new Regex(regexCode, RegexOptions.Singleline |
       RegexOptions.IgnoreCase);
       testo = regex1.Replace(testo, " ");

       textBox1.Text = testo;
       string[] words = testo.Split(' '); // words è un array di strings

       foreach (string wrd in words) // wrd takes each element in the array words
       {
              if (count.ContainsKey(wrd)
              {
                     count[wrd] = count[wrd] + 1;
              }
              else
              {
                     count[wrd] = 1; // se non la trovo, creo la entry = 1
              }
       }
}
```

Appendix 2. Programming Code Java– the Co-authorship Dynamic Network

In Chapter 3, to create both co-authorship and bibliographic coupling networks I used similar coding structures. The reported example has been used to generate yearly slices of a co-authorship network that is represented as a sparse graph in which the first column is the source vertex, the second is the target vertex, and the third is the year.

The input file is a .txt file generated by EndNote ® that has as lines the various articles and as a content many information, among which also the names of co-authors in the following fashion: Surname1, Name1, … , SurnameN, NameN.

The script adopts the Java sintax, but it is made to run in Processing. To run in Java, it needs minor adjustments.

```
// NETWORK OF COAUTHORS STARTING FROM THE ENDNOTE FILE //
ArrayList Autori = new ArrayList();
ArrayList sparseNetwork = new ArrayList();
ArrayList Yr = new ArrayList();
ArrayList temp; //here I store the total number of authors
int totalAuthors = 0;
PrintWriter output, out2;

void setup ()
{
  String nameFile = "inputFile.txt";
  int colAnno = 2;
  int colNomi = 1;
  String [] rows = parseLine(nameFile); //this reads the file and returns it line by line
  int [] vID = getVector("vectorID_noAbs_noWrongTopic_noRef.txt",0,"num");
  int matchAutori = 0;
  int numeroCoautori [] = new int [rows.length];
  int [] anno = new int [rows.length];
  for (int i = 0; i<rows.length; i++)
  {
    anno[i] = readYear(rows,i,colAnno);
  }

  int primoAnno = min(anno);
  int ultimoAnno = max(anno);
  int currentAnno= primoAnno;

  outerloop:
  for (int i =0; i<rows.length; i++)
  {
    if (anno[i] != currentAnno && anno[i] > primoAnno && anno[i]<ultimoAnno)
    {
```

```
      printStuff(numeroCoautori,primoAnno,currentAnno);
      currentAnno = anno[i];

    }

    if (anno[i] != currentAnno && currentAnno == 2012)
    {
      printStuff(numeroCoautori,primoAnno,currentAnno);
      break outerloop;
    }

    currentAnno = anno[i];
    if (anno[i] >= primoAnno && anno[i]<ultimoAnno && anno[i] == currentAnno)
    {
      //it takes the first column of the txt
      String nomi = takeNames(rows,i,colNomi);
      //it returns Lastname1,Name1,Lastname2,Name2,...
     String [] nomiInc = divideNames(nomi,i);
      int numAutori = countNames(nomiInc); //counts LastnameFN in the record
      numeroCoautori[i] = numAutori;

     //LastnameFN: BiscaroC, HuffAS, etc.
      String [] nomiAutori = fullNames(nomiInc, numAutori);
      int totalAu = contaAutori (nomiAutori);
      // Now I add authors to an arraylist
      matchAutori = compileNetwork(nomiAutori, i, matchAutori, vID,anno);
    }


  }
  printAutori();
}

// function that returns the year of publication of a paper
int readYear(String [] righe,int i, int colonna)
{
  String [] temp =split(righe[i],"\t");
  int tempInt = Integer.parseInt(temp[colonna]);
  return tempInt;
}

// function that matches co-authors of a paper in all other papers previously written
int compileNetwork(String[] nomiAutori, int i, int matchAutori, int [] vID, int[] anno)
{
  for (int au = 0; au <nomiAutori.length; au++) //short loop
  {

    int matching = -1;
    if(Autori.size() != 0)
    {
      outerloop:
      for (int j = 0; j<Autori.size();j++) //loop on the ArrayList  Autori j
      {
        if (nomiAutori[au].equals(Autori.get(j)))
        {
          matching = j;
          matchAutori = matchAutori +1;
          break outerloop;
        }
        else
        {
```

```
          matching = -2;
        }

      }
      if (matching == -2)
      {
        Autori.add(nomiAutori[au]);
        sparseNetwork.add(Autori.size());
        sparseNetwork.add(vID[i]);
        Yr.add(anno[i]);
        Yr.add(anno[i]);
      }
      if (matching > 0)
      {
        sparseNetwork.add(matching+1);
        sparseNetwork.add(vID[i]);
        Yr.add(anno[i]);
        Yr.add(anno[i]);
      }
   }
   if(Autori.size() == 0)
    {
      Autori.add(nomiAutori[au]);

     //it prints first the position of author in the ArrayList
      sparseNetwork.add(Autori.size());
      sparseNetwork.add(vID[i]);
      Yr.add(anno[i]);
      Yr.add(anno[i]); //then the paper plus one
    }
  }
  return matchAutori;
}

//get a vector from a txt file
int [] getVector (String nameFile, int header, String type)
{
  String row [] = loadStrings(nameFile); //here I put my file to load
  int [] vettore = new int [(row.length)-header];

  if (header == 0 && type.equals("num"))
  {
    for (int i = header; i<row.length; i++)
    {
      vettore[i-header] = Integer.parseInt(row[i]);
    }

  }
  return vettore ;
}

//function that returns Lastname and FirstName in one word: Warglien M -> WarglienM
String [] fullNames (String [] columns, int numeroAutori)
{
  String [] nomiCompleti = new String [numeroAutori];
  int position = 0;
  for (int c = 0; c< columns.length; c++)
  {
    columns[c] = trim(columns[c]);
    columns[c] = columns[c].replaceAll(" ",""); //I remove all white spaces
    int caratteri = columns[c].length();
```

```
      int conta = 0;
      for (char ca : columns[c].toCharArray())
      {
        if (Character.isUpperCase(ca))
        {
            conta = conta +1;
        }
        if (conta == caratteri)
        {
          numeroAutori = numeroAutori+1;
        }
        if (conta == caratteri)
        {

          if(columns.length == 1)
          {
            nomiCompleti[position] = columns[c];
          }

          if (columns.length != 1)
          {

            nomiCompleti[position] = columns[c-1] + columns[c].charAt(0);
            position = position +1;
          }
        }
      }
    }
    return nomiCompleti;
  }

  //function that prints the network and other data (like the entire list of coAuthors and
//number of coauthors for each paper)
  void printStuff(int [] numeroCoautori,int primoAnno, int currentAnno)
  {
    printNtwk(primoAnno,currentAnno);
    //printAutori();
    //printNumeroCoAutori(numeroCoautori);
  }

  void printNtwk(int primoAnno,int currentAnno)
  {
    //giving the name to the new file
    output = createWriter("NEWntwkCoau_YR_" + primoAnno + "_" + currentAnno +".txt");
    for (int i = 0; i< sparseNetwork.size(); i++) //
    {

      if (i%2==0) //check if it is EVEN
      {
        output.print(sparseNetwork.get(i) + "\t");
      }
      else
      {
        output.println(sparseNetwork.get(i) + "\t" + Yr.get(i));
      }
      //output.println(Yr.get(i));
    }

    output.flush();
    output.close();
  }
```

```
//printAutori prints the list of all co-Authors
void printAutori()
{
  output = createWriter("NEWnamesAuthors.txt");
  for (int i = 0; i< Autori.size(); i++)
  {
      output.println(Autori.get(i));
  }
  output.flush();
  output.close();
}


//printNumeroCoautori prints the number of co-authors of each paper
void printNumeroCoAutori(int [] numeroCoautori)
{
  output = createWriter("NEWnumberCoauthors.txt");
  for (int i = 0; i<numeroCoautori.length ; i++)
  {
      output.println(numeroCoautori[i]);
  }
  output.flush();
  output.close();
}

//Read Txt Function
String [] parseLine (String nameFile)
{
  String rows [] = loadStrings(nameFile); //here I put my file to load
  return rows ;
}

String takeNames (String [] righe,int i,int colonna)
{
  String [] names =split(righe[i],"\t");
  return names[colonna];
}


//this function splits a text according to given separator and produces a vector of
//strings
String [] divideNames (String firstColumn, int i)
{
  firstColumn = firstColumn.replaceAll(" and ",",");

  //this removes an error at O'Connor --> into OConnor
  firstColumn = firstColumn.replaceAll("\'","");
  String wordRegex = "\\b\\w+\\b";  //regex to match
  if (firstColumn.equals(""))
  {
    firstColumn = "MANYAUTHORS";
  }

  String [] columns = split (firstColumn, ",");
  String temp [] = match(columns[0],wordRegex);
  columns[0] = temp[0];

  return columns;
}
```

```
//This functions counts the co-authors of paper i
int countNames (String [] columns)
{
  int numeroAutori = 0;
  for (int c = 0; c< columns.length; c++)
  {
    columns[c] = trim(columns[c]);
    columns[c] = columns[c].replaceAll(" ",""); //removing all white spaces
    int caratteri = columns[c].length();
    int conta = 0;
    for (char ca : columns[c].toCharArray())
    {
      if (Character.isUpperCase(ca))
      {
          conta = conta +1;
      }
      if (conta == caratteri)
      {
        numeroAutori = numeroAutori+1;
      }
    }
  }
  return numeroAutori;
}


//This functions counts the total amount of different co-authors
int contaAutori(String [] nomiAutori)
{
  totalAuthors = totalAuthors + nomiAutori.length;
  println(totalAuthors);
  return totalAuthors;
}
```

Estratto per riassunto della tesi di dottorato

Studente: Claudio Biscaro

Matricola: 955666

Dottorato: Economia Aziendale

Ciclo: 24

Titolo della tesi: Wiring Knowledge. Metaphors, boundaries and scientific performance

Estratto:

La tesi analizza il fenomeno della combinazione della conoscenza che è chiave per l'innovazione e per la generazione di nuova conoscenza e per l'innovazione. La tesi è composta da tre capitoli. Nel primo capitolo, si è a studiato longitudinalmente il lavoro di un gruppo di scienziati e come usano le metafore per coordinare l'avanzamento della conoscenza sia disciplinare che multidisciplinare. Nel secondo capitolo, si individua una metodologia per identificare i concetti che fanno da ponte tra discipline diverse attraverso l'analisi delle reti combinata all'analisi testuale. Nel terzo capitolo, si studia come la rete di co-autori e la posizione dell'articolo nella letteratura incidono sull'impatto scientifico dell'articolo.

Abstract:

The topic of the dissertation is knowledge combination that is key for innovation and the generation of new knowledge and for innovation. The dissertation comprises three chapters. In the first chapter, we longitudinally study the activity of a group of scientists and how they use metaphors to coordinate both disciplinary and multidisciplinary knowledge creation. In the second chapter, we show a method to identify concepts that bridge different disciplines through the analysis of textual networks. In the third chapter, we study how the structure of co-authorship network and the position of the articles in the literature affect their scientific impact.

Firma dello studente