



Working Paper Series

SWPS 2014-10

June, 2014

Scientometric Mapping as a Strategic
Intelligence Tool for the Governance of
Emerging Technologies

Daniele Rotolo, Ismael Rafols,
Michael Hopkins, and Loet Leydesdorff

SPRU Working Paper Series

The SPRU Working Paper Series aims to accelerate the public availability of the research undertaken by SPRU-associated people of all categories, and exceptionally, other research that is of considerable interest within SPRU. It presents research results that in whole or part are suitable for submission to a refereed journal, to a sponsor, to a major conference or to the editor of a book. Our intention is to provide access to early copies of SPRU research.

Editors

Tommaso Ciarli
Daniele Rotolo

Contact

T.Ciarli@sussex.ac.uk
D.Rotolo@sussex.ac.uk

Associate Editors

Florian Kern

Area

Energy

F.Kern@sussex.ac.uk

Paul Nightingale

Science, & Technology Policy

P.Nightingale@sussex.ac.uk

Matias Ramirez

Development

Matias.Ramirez@sussex.ac.uk

Joe Tidd &
Carlos Sato

Technology Innovation Management

J.Tidd@sussex.ac.uk
C.E.Y.Sato@sussex.ac.uk

Maria Savona &
Mariana Mazzucato

Economics of Technological Change

M.Savona@sussex.ac.uk
M.Mazzucato@sussex.ac.uk

Andrew Stirling

Transitions

A.C.Stirling@sussex.ac.uk

Caitriona McLeish

Civil military interface

C.A.McLeish@sussex.ac.uk

Administrator

Jenny Lieu

J.Lieu@sussex.ac.uk

Disclaimer

The works available here are the responsibility of the individual author(s) and do not necessarily represent the views of other SPRU researchers. As matters of policy and practice, SPRU does not endorse individual research contributions.

Guidelines for authors

Papers shall be submitted in pdf or Word format. They should contain a title, an abstract, and keywords. Papers should be submitted to one of the Editors, who will process them and send them to the appropriate Associate Editor. Two members of SPRU will be asked to provide a short written review within three weeks. The revised versions of the paper, together with a reply to the reviewers, should be sent to the Associate Editor, who will propose to the Editors its publication on the series. When submitting the authors should indicate if the paper has already undergone peer-reviewing, in which case the Associate Editors may decide to skip internal review process.

Website

SWPS: www.sussex.ac.uk/spru/research/sewps

IDEAS: ideas.repec.org/s/sru/ssewps.html

Scientometric Mapping as a Strategic Intelligence Tool for the Governance of Emerging Technologies

Daniele Rotolo^{*1}, Ismael Rafols^{†2,1}, Michael Hopkins^{‡1}, and Loet Leydesdorff^{§3}

¹SPRU — Science and Technology Policy Research, University of Sussex

²Ingenio (CSIC-UPV), Universitat Politècnica de València

³Amsterdam School of Communication Research (ASCoR), University of Amsterdam

Working paper — Version: June 13, 2014

Abstract

How can scientometric mapping function as a tool of 'strategic intelligence' to aid the governance of emerging technologies? The present paper aims to address this question by focusing on a set of recently developed scientometric techniques, namely overlay mapping. We examine the potential these techniques have to inform, in a timely manner, analysts and decision-makers about relevant dynamics of technical emergence. We investigate the capability of overlay mapping in generating informed perspectives about emergence across three spaces: geographical, social, and cognitive. Our analysis relies on three empirical studies of emerging technologies in the biomedical domain: RNA interference (RNAi), Human Papilloma Virus (HPV) testing technologies for cervical cancer, and Thiopurine Methyltransferase (TPMT) genetic testing. The case-studies are analysed and mapped longitudinally by using publication and patent data. Results show the variety of 'intelligence' inputs overlay mapping can produce for the governance of emerging technologies. Overlay mapping also confers to the investigation of emergence flexibility and granularity in terms of adaptability to different sources of data and selection of the levels of the analysis, respectively. These features make possible the integration and comparison of results from different contexts and cases, thus providing possibilities for a potentially more 'distributed' strategic intelligence. The generated perspectives allow triangulation of findings, which is important given the complexity featuring in technical emergence and the limitations associated with the use of single scientometric approaches.

Keywords: scientometric overlay mappings; emerging technology; strategic intelligence; governance; case-study.

*Corresponding author: d.rotolo@sussex.ac.uk, Phone: +44 1273 872980

†i.rafols@ingenio.upv.es

‡m.m.hopkins@sussex.ac.uk

§loet@leydesdorff.net

1 Introduction

Emerging technologies have the potential to exert a long-term influence on a wide range of domains of the socio-economic system by transforming existing industries or creating new ones as well as changing knowledge production processes and related technical regimes (e.g. Corrocher et al., 2003; Cozzens et al., 2010; Day and Schoemaker, 2000; Martin, 1995). The emergence of novel technologies is not only associated with societal benefits, but also with unexpected adverse and destabilising changes (Collingridge, 1980). The governance of emerging technologies has therefore assumed an increasing relevance becoming priorities and part of the research agendas of many national governments (Hilgartner and Lewenstein, 2014).

Developing policies to stimulate emergence as well as to reduce the lack of control and risk of the unintended effects of emerging technologies on the socio-economic system is however a difficult activity. Uncertainties, ambiguity, and rapid dynamics feature in the emergence process of which directionality is the result of a variety of factors such as visions, goals, and expectations of the actors involved (e.g. Geels, 2002; Stirling and Scoones, 2009; Van Lente and Rip, 1998). Emergence in complex social systems as science and technology is reflective, i.e. actors are at the same time regulated by and regulating the emergence process (de Haan, 2006). The governance assumes an emergent character resulting from the complex interactions of these actors whereby the explicit attempts to shape arrangements are only one part of this process (e.g. Kooiman, 1993; Verbong and Geels, 2007). Governance is the result of both by intentional and un-intentional influences, *de facto* including factors and actors non directly or purposefully aiming at governing (Rip and Van Amerom, 2010). Lock-in and irreversibilities (David, 1985) and risk of technological alternative to be crowded out (Martin, 1995) add to this complexity.

The governing of emerging technologies therefore requires tools of 'strategic intelligence', i.e. tools that are able to provide 'intelligence' inputs to the policy-making process for the development of governance capable to cope with the uncertain, ambiguous, and rapid evolvments of technical emergence (Guston and Sarewitz, 2002; Kuhlmann et al., 1999). Conventional tools of strategic intelligence are the science and technology foresight, innovation policy evaluation, and technology assessment. Yet, despite the large variety of scientometric studies developing approach for the detection and the analysis of emerging technologies (e.g. Chen, 2006; Ohniwa et al., 2010; Small et al., 2014), the use of scientometrics for strategic intelligence in the context of emerging technologies is less explored.

The present paper aims to illustrate the extent to which a recently developed stream of user-driven and interactive scientometric mapping techniques, namely overlay mapping (Rafols et al., 2010), qualifies as a tool of strategic intelligence for the governance of emerging technologies. The key principle of these mapping approaches is to project data representing a focal subject area over a *basemap*, which represents the totality of contemporary activity in scientific and technological outputs at the considered time. The *overlay* can be constructed to represent the knowledge production of individuals, organisations, communities, or entire emerging fields in science and technology. Those techniques can be applied longitudinally to inform about the dynamics of emerging technologies across the geographical, social, and cognitive spaces of emergence as well as on combinations of those. For example, an organisation's publishing and patenting activities can be projected on maps of science or Google maps across different levels such as journals, patent classes, or research topics (e.g. Klavans and Boyack, 2009; Leydesdorff et al., 2014, 2013; Waltman and van Eck, 2012).

We conduct our empirical analysis on three illustrative case-studies of emerging technologies in the biomedical domain. The case-studies are used as 'vignette' to illustrate the 'intelligence' potential of the outputs overlay mapping approaches allow generating. The mapping on the selected case-studies is carried out since early phases of emergence and over decades of observation by using publication and patent data.

Results shows the different types of informed perspectives on emergence overlay mapping can generate with relative low efforts of analysts and decision-makers. These include urban areas producing the large amount of knowledge as well as the knowledge of higher future impact, geographical distribution of the key organisational actors in collaborative networks, and main scientific disciplines and technological domains involved in the emergence including dynamics of these. In addition, overlay mappings show flexibility and granularity in the analytical process. These tools are flexible in the sense that they are not constrained to the institutional boundaries of the databases where relevant data are maintained (Griliches, 1994). The informative perspectives on the emergence can be built by using same or similar search strings across multiple sources of data. Granularity is provided by the possibility to cross multiple levels of analysis of the emergence process in order to reveal macro- and micro-dynamics. Flexibility and granularity also favour the integration and comparison of results from different contexts, which in turn supports the development of a more 'distributed' strategic intelligence on the emergence in science and technology.

The paper is structured as follows. In the next section, we introduce the basic idea underlying overlay mapping approaches and the diverse possibilities of applications those techniques provide for the analysis of emerging technologies. We then introduce the three case-studies and data sources in Section 3. Results are presented in Section 4 and discussed in Section 5. Section 6 concludes the study.

2 Emerging technologies and overlay mapping

The scientometric community has made great efforts in the development of a number of techniques to trace the dynamics in science and technology domains since the seminal works by Price (1965), Small (1973), Garfield (1979), and Callon et al. (1983). Well established examples of these methodological approaches are co-citation analysis (e.g. Garfield, 1987; Small, 1973) bibliographic coupling (e.g. Kessler, 1963), and co-words analysis (e.g. Callon et al., 1983; Cambrosio et al., 2006).

The growing attention to emerging technologies and their transformative power (e.g. Corrocher et al., 2003; Cozzens et al., 2010; Day and Schoemaker, 2000; Martin, 1995) has also led scientometric scholars to focus their efforts on methodological approaches for the detection and mapping of emergence in science and technology (e.g. Chen, 2006; Ohniwa et al., 2010; Small et al., 2014). Those efforts, jointly with the increase of computational power, the improved performance of processing algorithms, the higher technical capabilities in exploiting databases in a more comprehensive manner, have spurred the development of a number of mapping techniques among which overlay mapping.

The overlay mapping approach allows, with relatively low efforts of analysts and decision-makers, rapid investigations across several dimensions and units of analysis of the emergence process. The basic idea underlying this approach is depicted in Figure 1. Data (mainly of publications and patents) gathered to represent an entity's activity in science and technology domains are used to construct an *overlay*. This is projected over a *basemap* representing both the totality of contemporary activity in scientific and technological outputs at the time, and the more stable structure of these in the background (e.g. Rafols et al., 2010). An entity may specifically refer to individuals, organisations, communities, or, especially for the focus of this paper, entire emerging fields in science and technology. Basemaps can be used to trace emergence across three different spaces, that are, the geographical, social, and cognitive spaces.

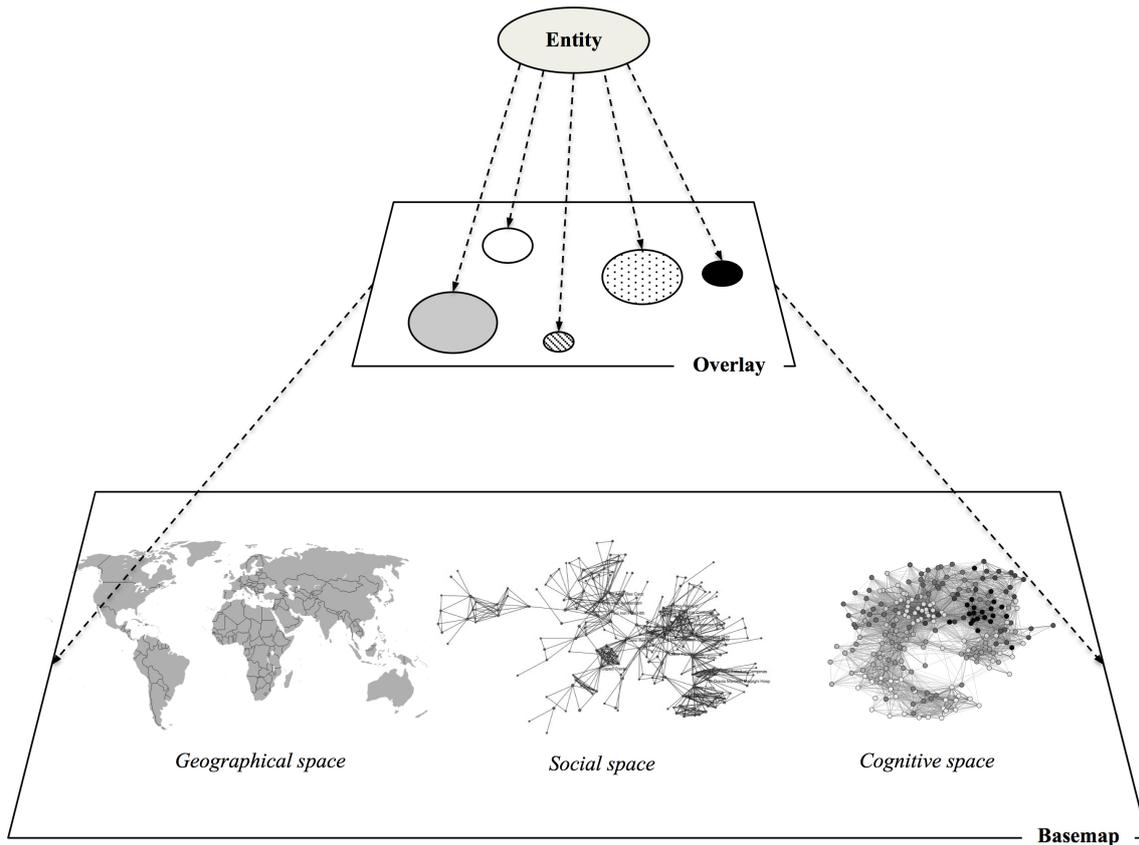


Figure 1: Overlay mappings: approach and spaces of emergence

These include geographical maps, inter-organisational networks, maps of science representing the entire structure of science at multiple levels of analysis such as disciplines, journals or topics, and maps of technological areas as identified by patent classes. These spaces of emergence were initially specified as the main dimensions in early scientometric projects (e.g. Narin, 1976; Small and Garfield, 1985) and related initiatives led by researchers in science and technology innovation studies (e.g. Irvine and Martin, 1985) and patent analysis (e.g. Jaffe and Trajtenberg, 2002).

We argue that overlay mapping can complement the existing set of more conventional tools of strategic intelligence for the policy-making of emerging technologies. These mapping techniques can generate evolutionary perspectives to potentially reveal intentional and unintentional government arrangements among factors and actors involved in the emergence process. Those informative perspectives may historically and timely inform analysts and decision-makers on relevant dynamics of emergence such as main geographical areas involved in the knowledge production processes and collaborative interactions among those, main actors in inter-organisational networks and their geographical distribution, domains of science and technologies involved in the emergence and associated dynamics across disciplines, journals, technological classes, and (in the case of biomedical technologies) medical areas. The increased knowledge on emergence

overlay mapping may provide the policy-making process with can in turn support the development of forms of governance that are more capable to cope with the uncertainty, ambiguity, and rapid dynamics of emergence.

Flexibility in the use of diverse sources of data and granularity in the multiple levels of analysis are additional features that qualify overlay mappings as a tool for a 'distributed' strategic intelligence on emerging technologies. Results of the analyses and associated knowledge from different contexts can indeed reflexively and discursively be integrated and compared with relative low efforts. The choice of overlays and basemaps as well as combinations of them however depends on the specific question(s) the analyst or decision-maker aims to address.

To examine the use of overlay mapping as tools of strategic intelligence for the governance of emerge in science and technology, we first summarise and then apply a number of overlay mapping techniques, that have been developed and reviewed in the domain of the information sciences and scientometrics, on three case-studies of emerging technologies. Methodologically, we rely on previous works some of the authors of this paper have conducted. We specifically build on (i) Leydesdorff and Persson (2010) for mapping co-authorship relations in publications as overlay to Google Maps, (ii) Rafols et al. (2010) for mapping publications in terms of Web of Science (WoS) Categories, (iii) Bornmann and Leydesdorff (2011) for the mapping of excellence in publications, (iv) Leydesdorff and Bornmann (2012) for the equivalent mapping of patents, and (v) Leydesdorff et al. (2012) for mapping medical innovations in terms of Medical Subject Headings (MeSH).¹

3 Methods

3.1 Background on the case-studies

We conduct our analysis on three-case-studies of emerging technologies in the biomedical domain: (i) RNA interference (RNAi), (ii) cervical cancer testing technologies for Human Papilloma Virus (HPV) and (iii) genetic testing for the Thiopurine Methyltransferase (TPMT) enzyme. These case-studies were selected since they exhibit diversity in terms of their contexts and positions in the innovation chain. From a governance perspective, each technology raises different challenges:

¹ Reiterating the full technical details of these methods goes beyond the scope of this paper. The reader is referred to the relative studies and corresponding webpages. In addition, it is worth noting that these methods build on a broader set of contributions on S&T mapping for policy use (e.g. Börner, 2010; Noyons, 2005, 2001).

RNAi is a new technology with radical potential across the domains of therapeutics, diagnostics and as a basic research, but with fluctuating industrial interest; HPV testing is battling the entrenched technology of Pap testing, with powerful interest groups on both sides; TPMT testing technologies are exploited in a series of small clinical niches, mainly in developed countries where there are concerns that best practice may not be spreading. This variety provides us with the opportunity to enrich our discussion the use of overlay mapping to generate 'intelligence' inputs for the governance of emerging technologies. In the followings, we provide background sketches for the case-studies to contextualise the scientometric maps presented later in the paper.

Firstly, RNAi is a molecular process that can silence the expression of genes. Genes play a critical role in the progression of cancers, genetic diseases, and infection agents. Theoretically, by silencing specific genes one can stop the progression of a given disease. This 'small RNA' silencing mechanism was discovered in 1998.² Its discovery reshaped the landscape of research on gene expression creating important expectations especially for the therapeutic applications (Sung and Hopkins, 2006). RNAi can be conceived as a general purpose technology for research in labs.

Secondly, HPV testing technologies are positioned within a specific domain of application, i.e. the detection for HPV virus, which is recognised as causing cervical cancer.³ Cervical cancer has a significant disease burden — about 500,000 new cervical cancers occur and cause about 250,000 deaths worldwide each year. This has led to the development of a large screening program, especially in developed countries.⁴ In the US more than 100 million tests are performed annually.⁵ While cervical cancer screening has been conducted for more than 50 years using the (cytology-based) Pap test to detect cancerous cells or cells potentially evolving in cancerous states, the discovery of the strong association between HPV infections (especially HPV types 16 and 18) and cervical cancer in the 1980s opened the space for the development of a competing and more sensitive testing technology based on molecular biology (Hogarth et al., 2012).

Thirdly, similarly to HPV test, TPMT testing technologies are positioned close to the

² Andrew Z. Fire and Craig C. Mello were awarded the 2006 Nobel Prize in Physiology or Medicine 2006 for this discovery (Fire et al., 1998).

³ Harald zur Hausen 1987, who discovered the association between the HPV and cervical cancer, was awarded the 2008 Nobel Prize in Physiology or Medicine.

⁴ Cervical cancer remains however one of the most common cancers among women in developing countries that account for about 85% and 88% of new cases and deaths, respectively (GLOBOCAN 2008 available at <http://globocan.iarc.fr>).

⁵ Health United States 2012 — Technical report.

applied-research domain. Yet, its application for clinical utility is contested across medical fields (e.g. different clinical guidelines supporting and discouraging the use of the test). TPMT testing is one of an emerging class of 'pharmacogenetic tests' which predict adverse events associated with pharmaceutical use (Hopkins et al., 2006). TPMT is an enzyme in the human body responsible for metabolising (i.e. breaking down) thiopurine drugs. Cytotoxic thiopurine drugs such as "Azathioprine" are used to treat a range of conditions including leukaemia, and autoimmune diseases (such as Lupus, or rheumatoid arthritis). However, when a patient has mutations in the gene encoding TPMT, she/he is unable to metabolise the drug and is at increased risk of toxicity from a build-up of thiopurines. Therefore, several types of TPMT testing technologies have begun to emerge with applications across a number of clinical fields of use such as oncology, dermatology, gastroenterology, and rheumatology.

3.2 Data sources

Data were collected from multiple databases. We relied on Thomson-Reuter's Web-of-Science (WoS) and MEDLINE/PubMed for publication data, and on the United States Patent and Trademark Office (USPTO) for patent data. These databases were queried with a set of keywords we identified by triangulating interviews with experts and quantitative research works on the case-studies (see Table 1). The interviews with experts in the fields of the selected case-studies are also used as lens of interpretations of the results we obtain from the application of overlay mapping approaches.

The retrieval of data for emerging technologies is challenging. Research topics often overlap and the vocabulary they use is in flux. Recently developed methods for retrieval rely on combining lexical search and clustering citation methods (Zitt and Bassecoulard, 2006), yet these approaches would be time-consuming for the analyst. We therefore deemed more suitable to use keyword-based searches given the parsimony and timeliness of the approach. We limited the search of keywords and their combinations in scientific articles' titles since this approach tends to generate less false positives compared to the searching in articles' abstracts. Abstracts indeed often contain technical and methodological terms not representing the core of knowledge the given article claims (Leydesdorff, 1989). An extended search may generally retrieve many additional records, yet with the risk of including many ones not closely related to the given emerging technology one aims to trace — it increases 'recall' at the expenses of losing 'precision', in bibliometric technical terms.

Table 1: Data sources and search strings.

Case-study	Data	Database	Search string
RNAi	Publications	ISI WoS	TI=siRNA or TI=RNAi or TI="RNA interference" or TI="interference RNA"
		MEDLINE/PubMed	siRNA[Title] or RNAi[Title] or "RNA interference"[Title] or "interference RNA" [Title]
	Patents	USPTO	ACLM/(siRNA or RNAi or "RNA interference" or "interference RNA")
HPV testing	Publications	ISI WoS	(TI=HPV* or TI="Human Papilloma Virus*" or TI="Human Papillomavirus*" or TI="Human Papilloma*virus*") and (TI=Cervical or TI=Cervix) and (TI=diagnos* or TI=test* or TI=assay or TI=detect* or TI=screen* or TI=predict*)
		MEDLINE/PubMed	(HPV*[Title] or "Human Papilloma Virus*" [Title] or "Human Papillomavirus*" [Title]) and (Cervical[Title] or Cervix[Title]) and (diagnos*[Title] or test*[Title] or assay[Title] or detect*[Title] or screen*[Title] or predict*[Title])
	Patents	USPTO	ACLM/((HPV or "Human Papilloma Virus\$" or "Human Papillomavirus\$") and (Cervical or Cervix) and (diagnos\$ or test\$ or assay or detect\$ or screen\$ or predict\$))
TPMT testing	Publications	ISI WoS	TI=TPMT or TI= "Thiopurine Methyltransferase"
		MEDLINE/PubMed	TPMT[Title] or "Thiopurine Methyltransferase"[Title]
	Patents	USPTO	ACLM/(TPMT or "Thiopurine Methyltransferase")

The identification of patents that are relevant for the building of informative perspectives on emerging technologies requires a different approach. The incentives to patent are indeed different from those underlying the publication of scientific articles (Murray, 2002). The primary purpose of the patent system is to reward patentees by providing them a temporary monopoly to commercially exploit the patented inventions. Yet, this requires patentees to disclose the technical knowledge of the inventions. In this regard, patent attorneys are very careful in including valuable information in the appropriate sections of patents. Among these sections, claims are the most relevant source informing on the scope of the technical knowledge (Hunt et al., 2007). Claims "define the invention and are what aspects are legally enforceable" (USPTO Glossary).⁶ We therefore focus the search of keywords on the claims in patents. Issues related to the definition and delineation of the boundaries of emerging technologies are further discussed in the next section.

The number of publications and patents that relate to each of the three cases, from 1982 to 2011, are reported in Table 2. While the first publications related to TPMT and HPV stem from the early 1980s, data for RNAi appears only since 1998 when this silencing mechanism was discovered and first published. Publication data from WoS and MEDLINE/PubMed allow for relatively simple but informative reports on the rapid emergence of these three technologies in terms of published scientific articles. The fast growth provide evidence of the emerging character of the selected technologies (Cozzens et al., 2010). It is clear that the pace of this growth as well as the scale of this emergence is significantly different from one case to another in two respects. First, the growth in the number of publications for RNAi is steeper than the other two case-studies. Second, RNAi and HPV testing technology show an increasing number of publications for the entire observation period. Conversely, the testing technology for TPMT enzyme seems to have reached its mature phase in the last few years of observations.

Patent data reveal similar distinctive features. The production of patents (both granted patents and patent applications) related to RNAi, for example, is relatively much greater than in the case of HPV and TPMT testing technologies. This is not surprising given the former's wide potential of applications. The declining of the patenting activity related to RNAi in the last two years of observation is mainly due to the USPTO publishing patent applications after 18 months from the filing date. Although most patenting in RNAi occurs in small firms, the

⁶ The USPTO Glossary is available at <http://www.uspto.gov/main/glossary>.

Table 2: Publication and patent data for the three case-studies.

Year	RNAi			HPV testing			TPMT testing		
	ISI WoS	MEDLINE PubMed	USPTO (granted)	ISI WoS	MEDLINE PubMed	USPTO (granted)	ISI WoS	MEDLINE PubMed	USPTO (granted)
1982				0	0	0	3	2	0
1983				2	1	0	2	0	0
1984				0	0	0	0	2	0
1985				4	2	0	3	2	0
1986				12	10	0	5	1	0
1987				22	14	4	5	4	0
1988				17	16	0	2	0	0
1989				30	24	3	1	1	0
1990				32	28	0	2	2	0
1991				36	31	1	6	5	0
1992				41	49	2	5	5	0
1993				18	27	1	11	9	0
1994				29	30	4	8	5	1
1995				24	28	5	17	14	1
1996				35	32	1	11	8	0
1997				33	28	2	15	10	0
1998	3	2	1	37	33	1	20	15	1
1999	15	10	1	41	35	4	21	11	0
2000	42	33	4	43	41	4	24	15	0
2001	56	52	5	50	52	1	28	19	3
2002	166	126	25	54	43	2	43	25	1
2003	427	280	45	69	59	9	42	30	1
2004	785	523	83	59	59	8	36	23	1
2005	892	681	86	105	95	6	40	19	1
2006	932	782	124	94	86	3	39	28	2
2007	1002	737	131	121	91	6	30	15	0
2008	1032	827	106	134	106	5	26	15	4
2009	1016	837	118	154	120	5	43	30	0
2010	1029	904	89	121	99	0	24	19	0
2011	1131	969	35	143	118	0	28	16	0
Total	8528	6763	853	1560	1357	77	540	350	16
			5951			210			41

Notes. Data were collected in January 2013. Data on USPTO patent applications are accessible since year 2001. The filing year of patents was considered.

decision of some large pharmaceutical companies, including Merck, Roche and Pfizer, to shut down their R&D units on RNAi may have also resulted in a decreased interest on RNAi and contributed to this decline (Lundin, 2011).

The patent application activity around HPV testing technology grows from 2002 to 2004 and then stabilises in the subsequent years with a peak of applications in 2009. The low number of granted patents and patent applications for TPMT testing technology does not allow clear trends in the production of technical knowledge to be deduced, other than to note this is a less vigorous field.

3.3 Definition and delineations issues

A key preliminary input for the use of mapping and overlay techniques is the identification of the boundary of emerging technologies. In general the definition of a technology is more ambiguous than one may initially expect. An emergent technology can be defined as the knowledge underpinning method or processes to fulfil a purpose (e.g. understanding of gene silencing) or as a network of practices and components (e.g. the molecules and techniques that result in gene silencing). The case of RNAi is apt to highlight the ambiguity inherent in the delineation of the boundary of emerging technologies.

RNAi is the naturally occurring process in which gene expression is reduced as a result of the break down of messenger RNAs. This process can be triggered by naturally occurring (endogenous) molecules, called microRNAs (miRNAs), or by externally inserted (exogenous) molecules, called small interfering RNAs (siRNAs). Soon after the discovery of RNA interference by Fire et al. (1998), siRNAs were recognised as valuable for gene silencing, both as tools for research and for therapeutic purposes, which lead to a boom in public and private R&D investments (Haussecker, 2008, 2012). In parallel, but with some significant delay, it was realised that miRNAs were not a marginal phenomenon, but played a major role in gene regulation, including abnormal down or up regulation in certain diseases caused or earmarked by anomalous gene expression, such as many cancers.

In principle, then, research on the science and technology of RNAi should include both siRNAs and miRNAs. However, it turns out that 'RNA interference' as a technology (not as phenomenon) has become mainly known as the human-induced, exogenous interference, which was developed for therapeutic purposes. We then have two potential definitions of RNAi, one covering the entire field and one excluding miRNA — commonly adopted for the discussion of

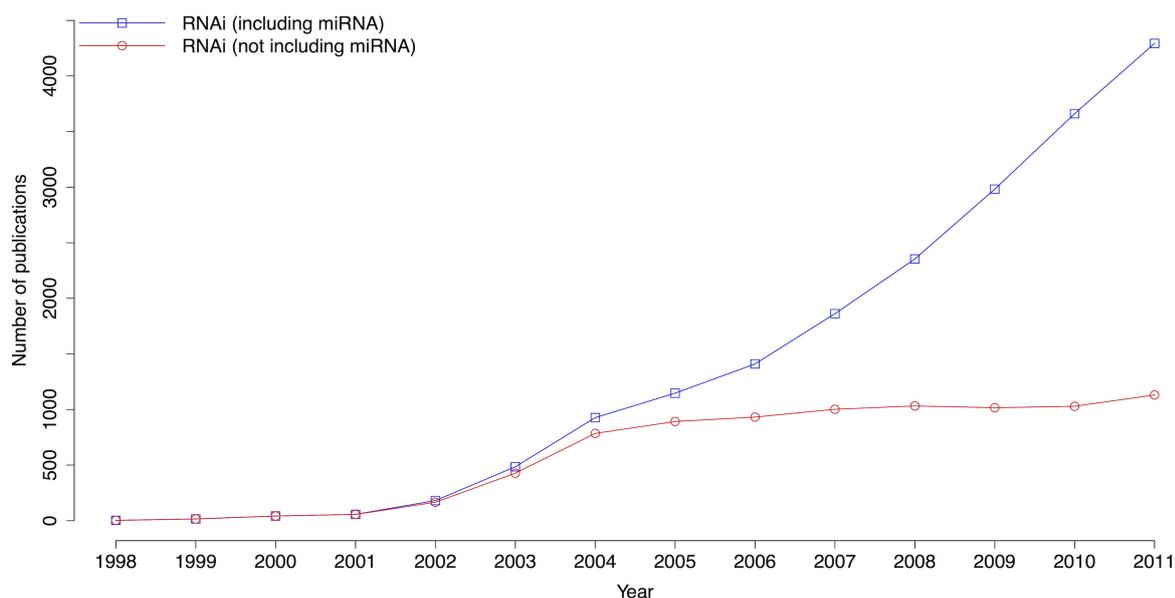


Figure 2: Delineating the boundary of RNAi. Scientific articles related to miRNA were retrieved by using the following search string in WoS: 'TI=microRNA* or TI=miRNA*'.

therapeutic applications.

Figure 2 illustrates that the observed dynamics are different. Whereas miRNA research is still booming, possibly due to their potential use as biomarkers in diseases such as cancers, publications focused on siRNA have reached a plateau of about a thousand publications per year. The latter is possibly related to the challenges encountered in delivering siRNA in therapeutic applications that resulted in a retreat of pharmaceutical investment (Haussecker, 2012). These differences in trends support the views that the use of miRNAs and siRNAs follow distinct trajectories, which it is best to differentiate. In this study, we focus on RNAi as research tool and for therapeutic applications, i.e. not including miRNA.

4 Results

The use of overlay mapping approaches as tools of strategic intelligence for the policy-making process is discussed below. We specifically build perspectives on the emergence process of the aforementioned three illustrative case-studies across the geographical, social, and cognitive spaces. Due to space limitations, we discuss and report in the paper a sample of the results obtained — the entire set of maps and overlays is available as supplementary materials, which include interactive features, at www.interdisciplinaryscience.net/overlaymapping.

4.1 Tracing emergence in the geographical space

One can combine the use of overlays with geographical (base)maps to visualise, by using publications and patent data, the emergence process across cities, regions, and nations. Building on previous works (e.g. Bornmann and Leydesdorff, 2011; Bornmann and Waltman, 2011; Hu et al., 2012; Leydesdorff and Bornmann, 2012), one can, for example, identify sites for a given emerging technology where highly cited scientific articles were published more frequently than expected. This is depicted in terms of the sizes and colours of nodes.⁷

For the three case-studies of this article, we used overlays projecting publication data according to a 5-year time window.⁸ Using constant time windows across the case-studies allows us capturing the different paces at which the diverse selected emerging technologies have emerged in terms of (WoS) publications. Figure 3 depicts the results applied to HPV and TPMT testing technologies during the period 2002-2006.

For the case of HPV testing technologies, these maps identify, over the entire observation period, European urban areas producing highly cited scientific knowledge more frequently than expected in London, Paris, and Amsterdam. New areas have also started to appear both in the North (nearby Copenhagen, Helsinki, and Jena) and South (nearby Barcelona, Bologna, and Turin) of Europe since the mid-1990s. The US sites contributing to this emerging technology with highly cited articles are mainly located on the coasts, specifically in the area of Washington D.C., Baltimore, New York, and Boston, for the East Coast, and near San Francisco and San Diego for the West Coast. Georgetown University and the private company Digene in the area of Washington D.C. have played a key role for the development and the adoption of a test for the HPV detection in cervical cancer screening (Hogarth et al., 2012). The maps also reveal the rise of new sites producing highly cited scientific knowledge in South America (e.g. near Sao Paulo, Buenos Aires) during the last ten years of observation.

This geographical mapping for TPMT testing technology locates, at the beginning of the observation period, highly cited articles in the urban areas of Rochester and Memphis (US) as well as Sheffield (UK). However, as for HPV testing technology, new sites have started to appear

⁷ The z - test for two independent proportions is used. The null hypothesis is the randomness in the selection of papers for a city (see Bornmann and Leydesdorff, 2011). A threshold of top 10% most-frequently cited scientific articles is selected.

⁸ Similar dynamics are observed by using narrower/broader time windows (e.g. 3-year, 7-year). It is worth noting that in order to exploit all the available data up to year 2011 the first time window for RNAi is 1998-2001, i.e. four years of observations. RNAi can indeed be observed starting from year 1998 when this gene silencing mechanism was discovered (Fire et al., 1998).

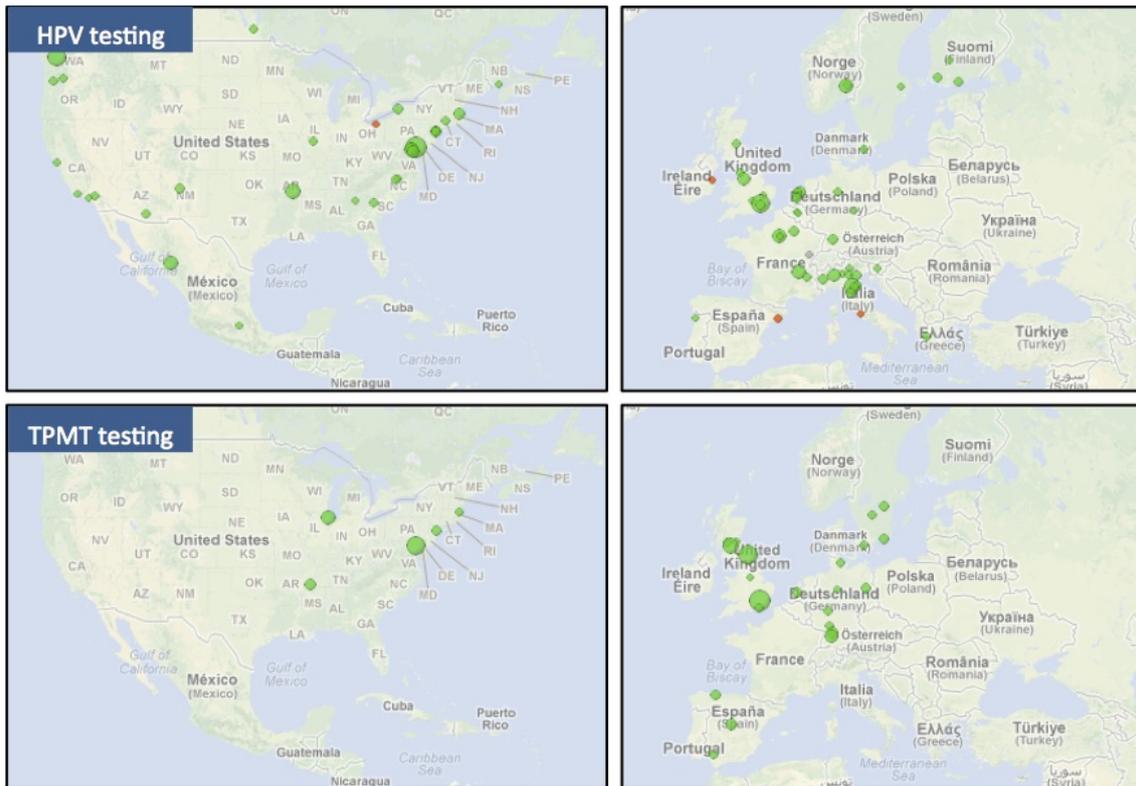


Figure 3: Areas of highly cited publications for the 2002-2006 period. Nodes are coloured dark green (red) when the difference between the observed number of top-cited publications and the expected one is positive (negative) and statistically significant ($p < 0.05$), light green (orange) otherwise.

since 1997 close to the East and West coasts (e.g. areas of Washington D.C., Boston, San Diego, San Francisco), across the UK (e.g. near London, Glasgow, Edinburgh) and Europe (e.g. Berlin, Madrid, Seville).

In the case of RNAi, the mapping shows the emergence of urban areas producing highly cited articles on the East Coast of the US near Baltimore and New York, and, especially, in the area of Boston (see supplementary materials). The RNAi mechanisms were discovered by groups of scientists working in Massachusetts (Haussecker, 2008). However, while the maps in the subsequently years report the emergence of many sites all round the world, Massachusetts has persisted as the main area producing highly cited scientific articles — this is revealed by the size of the node and as one of the hubs for RNAi biotech firms.

Interesting dynamics can be observed by combing these findings with those obtained using a similar approach but geo-localising patent data.⁹ These maps are reported for the 2002-2011 period in Figure 4 — the first time window (1998-2001) is excluded because there are only

⁹ For this approach, we decreased the threshold from top 10% (as used for publication data) to the top 25% cited patents, necessary because the number of publications is an order of magnitude higher than the number of patents (Leydesdorff and Bornmann, 2012).

11 patents, which constitute a too small sample for the statistical analysis.¹⁰ While the area of Denver has emerged as one of the many urban areas of high citation for the production of scientific knowledge in the later period of observation, these maps show this area producing highly cited patents more frequently than expected. These patents related to the reagents used in RNAi (small interfering RNAs) while patents produced in centres nearby New York and Philadelphia related to therapeutic applications. The analysis does not identify any urban area of high citations in Europe or Asia.

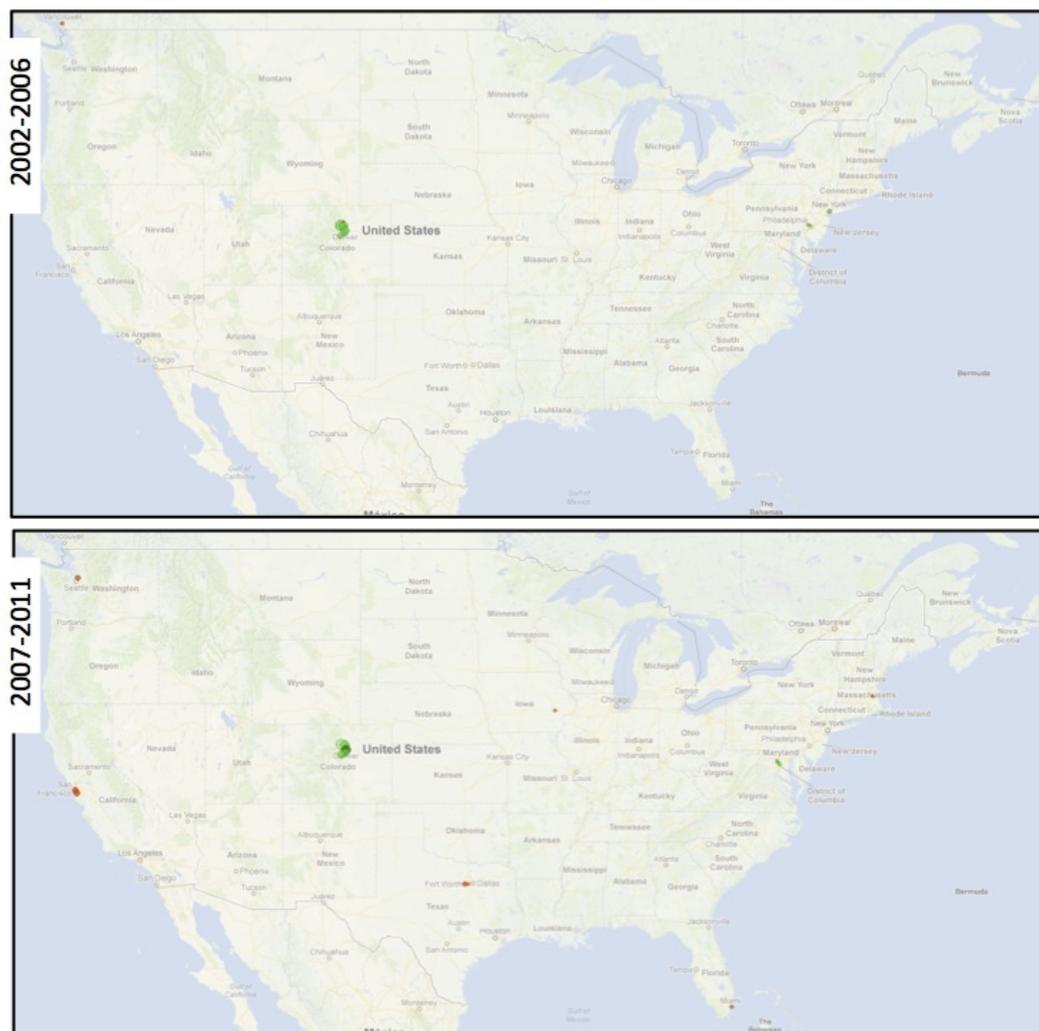


Figure 4: Areas of excellence (patent-based) for RNAi for the 2002-2011 period. Nodes are coloured dark green (red) when the difference between the observed number of top-cited patents and the expected one is positive (negative) and statistically significant ($p < 0.05$), light green (orange) otherwise.

This analysis shows how overlay mappings can provide the policy-making process an analyst is undertaking with relevant inputs such as the list of geographical areas where main advance-

¹⁰For this reason, these maps cannot be built for the HPV and TPMT testing technologies.

ments for a given emerging technology were achieved and areas that are 'unexpectedly' listed. It may also inform on areas that persistently contribute over time to the development of the emerging technology as well as on new emerging ones. *De facto* coordination arrangements are also revealed. For example, the maps for HPV and TPMT testing technologies identify a number of urban areas in which highly cited publications concentrate as located developing countries (see supplementary materials). A further analysis on the collaboration networks (see below) revealed co-authorship links between these areas and others leading the advancements of these emerging technologies in the developed countries. These collaborations may have provided developing countries with the access to critical capabilities and resources to produce novel and high-quality knowhow, but perhaps indicating some crucial contribution (e.g. a genetic resource or patient population) needed by researchers in developed countries too that was helpful.

It is important to note some limitations of the above-discussed approaches. Firstly, the geographical information reported in publication and patent data may not reflect the locations where the research was conducted. This is especially true when patent data are considered. Inventors list their home addresses which may not necessarily reflect the geographical locations where inventions were developed. Secondly, while the overlays built at city-level provide granularity, they can represent sites located in the same urban area as two different nodes. For example, in the case of HPV testing technology, Silver Spring (a suburb of the US capital) was considered as a node different from Washington D.C.. These limitations represent an important venue for future development of those techniques.

4.2 Tracing emergence in the social space

The structure of the relationships among the actors surrounding emerging technologies and their dynamics play a critical role in the emergence process (e.g. Latour, 1993; Powell, 1990). These connections are channels through which actors gain access to and mobilise knowledge, resources, and power. Networks of agents affect and are affected by emerging technologies, i.e. actors creates social structures over the emergence that both enable and constrain their actions (Giddens, 1984). By using co-authorship data (e.g. Crane, 1972; Glänzel and Schubert, 2004; Wagner, 2008), the dynamics across this relevant space of emergence can be traced. Novel techniques allow specifically building perspectives crosscutting both the social and geographical spaces (Leydesdorff and Rafols, 2011).

For example, Figure 5 shows the 2002-2006 collaboration networks at city-level for HPV and

TPMT testing technologies. In this map, nodes are cities and the linkages between nodes are traced by using co-authorship data.¹¹ Investigating the dynamics across these maps may provide informative perspectives that are derived by combining the geographical and social spaces. Examples of key empirical questions the overlay mapping approach may help to address in the policy-making process are: Where does an emerging technology arise? Does the collaboration network cluster in specific areas? Does the given emerging technology spread across cities, regions, and countries and, if yes, through which (collaboration) channels — which types of grant were important, which size of institutions, or types of knowledge brokers?

The collaborative network of HPV testing technologies, for example, discloses three relevant dynamics of the emergence process. First, a strong collaborative activity between the US (especially the areas of Washington and New York) and Europe (initially Germany) can be observed. This technology indeed started to emerge after a scientist, Harald zur Hausen, at German Cancer Research Centre proved HPV infections to be strongly associated with the development of the cervical cancer (zur Hausen, 1987). This discovery subsequently found important application in the US where extensive screening programs on cervical cancer were already in place. Here, a small biotech company, namely Digene, developed and marketed a series of FDA-approved HPV tests, which are recognised as significant milestones in the management of cervical cancer. Second, the last ten years of observation show an increasing involvement of developing countries (e.g. Brazil, India) in the research networks. Cervical cancer in these countries is a significant social burden. HPV is sexually transmitted and the costs associated with the screening of the population are not always affordable. Third, globalisation of research on HPV testing technologies can be observed across the entire period as revealed by the density of the network of relationships across cities. The latter dynamic is also observed in the case of RNAi that diffuses all round the world, even more rapidly than HPV and TPMT testing technologies (see supplementary materials).

In the case of TPMT testing, further informative dynamics are revealed. The co-authorship network evolves starting from a strong collaboration between Rochester (US) and Sheffield (UK). This collaboration intensifies over the entire observation period while major collaborative networks within the UK and US national boundaries formed later, only since the 1990s.

¹¹ Similar maps can be also built with the top-cited publications approach we described in the previous section. This provides additional perspectives on the structural position cities producing highly cited knowledge occupy in the web of collaborative relationships (co-authorships). We made these maps available for the three case-studies as supplementary materials.

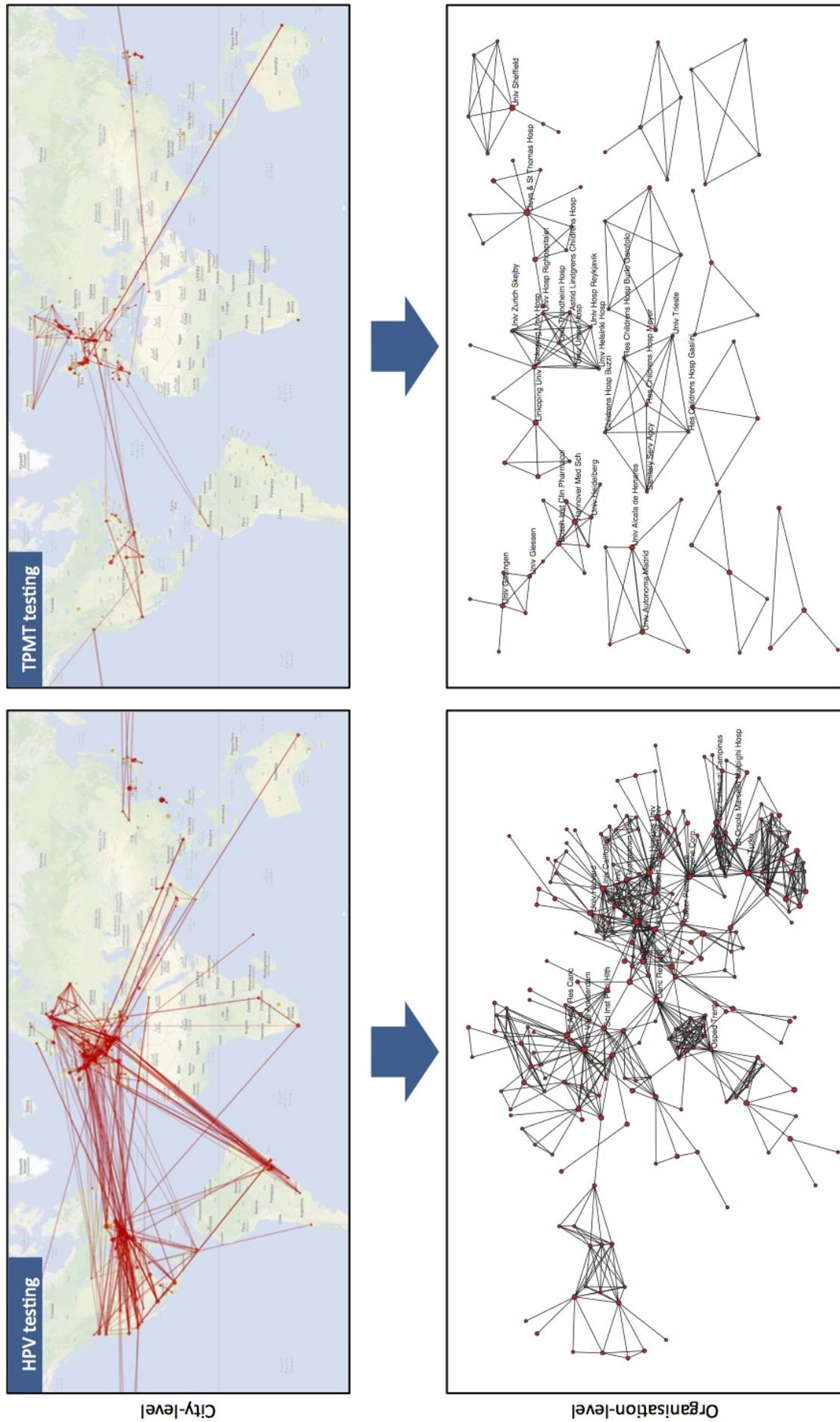


Figure 5: Crosscut on the geographical and social spaces of the emergence process for the 2002-2006 period. The size of nodes is proportional to the \log_2 of the number of scientific articles (plus one) that the given city (above) / organisation (below) published. We reported labels for the top 5% central organisations in the co-authorship network according to the degree centrality.

Subsequently, from 1997, we observed also the rise of the European network initially involving Germany, France, The Netherlands, and UK, and then including other countries such as Italy and Spain.

Building on these crosscuttings of the geographical and social spaces, one can focus more attention on the social dynamics by looking at the structure of the web of relationships composing the network at a lower level of analysis such as the organisation-level. The network can be explored with algorithms that identify cohesive groups of organisations as well as public and private players occupying key positions. The network analysis provides a broad range of measures for this purpose (Wassermann and Faust, 1994).

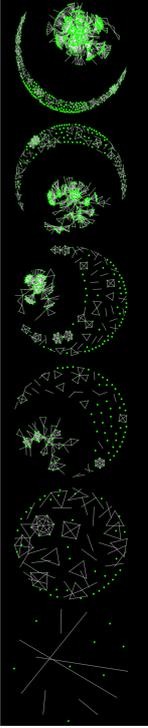
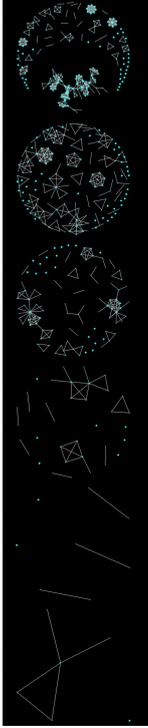
For example, the lower part of Figure 5 depicts the organisational collaborative networks corresponding to the aforementioned collaborative networks overlaid on the geographical space.¹² While for HPV testing technologies a giant component of the network can be identified, the organisational network for TPMT testing is highly fragmented, as revealed by the different separated groups of organisations (components), until the last five years of observation (Table 3). For clarity of the representation, Figure 5 depicts the largest component for HPV testing technology for 2002-2006 period. We instead represented for TPMT testing technology components formed by at least three organisations.

The dynamics of the collaborative network surrounding the research activity on HPV testing technology show Digene occupying a strong and influential position within this network by collaborating with major institutions in the field of cervical cancer screening (e.g. National Cancer Institute, a public sector sponsor, and Kaiser Permanente, a large healthcare service provider and early test adopter). This eventually allowed Digene to influence the adoption of the test, for example by working with clinicians involved in the definition of medical guidelines (Hogarth et al., 2012). In other words, while Digene's activity was 'regulated' by the FDA for example, Digene was affecting the developments and dynamics in cervical cancer screening, too.

As discussed, the co-authorship network for TPMT testing technology is characterized by the presence of several separate components over a significant part of the observation period.

¹² Organisations' names included in publication data of WoS present a number of variations, i.e. the same organisation may be spelled in different manners. We use The Vantage Point software to clean the data. This software specifically analyses and suggests groups of names that may refer to the same organisation by using a fuzzy algorithm that exploits also the information included in other fields of the publication data. We checked those suggestions for our sample of publications and confirmed those matches for which the manual desktop search over the Internet provided further support. Freeware routines for using institutional addresses but without this cleaning process can be retrieved at www.leydesdorff.net/maps.

Table 3: Structural properties of the co-authorship network at organisation-level.

	Time window		1982-1986	1987-1991	1992-1996	1997-2001	2002-2006	2007-2011
HPV testing technology	Nodes		20	130	173	265	471	816
	Ties		6	88	223	476	980	2075
	Density		0.03	0.01	0.01	0.01	0.01	0.01
	Avarage Path Length		1	1.3	3.4	2.8	4.4	4.4
	Isolated (%)		8 (40%)	35 (27%)	43 (25%)	47 (18%)	55 (12%)	83 (10%)
	Components		0	15	11	16	23	40
	First largest component (%)		2 (10%)	7 (5%)	62 (36%)	83 (31%)	239 (51%)	504 (62%)
	Second largest component (%)		2 (10%)	6 (5%)	8 (5%)	12 (5%)	14 (3%)	9 (1%)
	Degree, Mean (Std.Dev.)		0.6 (0.5)	1.3 (1.3)	2.6 (3.0)	3.6 (4.8)	4.2 (4.7)	5.1 (5.7)
								
TPTM testing technology	Nodes		6	8	36	111	200	232
	Ties		5	3	28	123	203	413
	Density		0.33	0.11	0.04	0.02	0.01	0.02
	Avarage Path Length		1.5	1	1.6	2.4	1.7	3.9
	Isolated (%)		1 (17%)	2 (25%)	7 (19%)	19 (17%)	43 (22%)	33 (14%)
	Components		1	0	3	11	22	19
	First largest component (%)		5 (83%)	2 (25%)	9 (25%)	25 (23%)	15 (8%)	82 (35%)
	Second largest component (%)		1 (17%)	2 (25%)	5 (14%)	17 (15%)	14 (7%)	9 (4%)
	Degree, Mean (Std.Dev.)		1.7 (1.4)	0.7 (0.5)	1.6 (1.4)	2.2 (2.2)	2.0 (2.0)	3.6 (3.2)
								

Note. The minimum size of a component is set to three nodes.

The network seems to initially develop around the strong collaboration between Mayo Clinic (Rochester, US) and University of Sheffield and subsequently around three actors, who played a key role for the creation of a large network component in the later years of observation. These actors are St. Jude Children’s Research Hospital, University of Manchester, and Dr Margarete Fischer Bosch Institute. Further analysis of the content of publications and patent data revealed St. Jude Children’s Research Hospital as conducting fundamental basic research in this domain as well as developing strong IP by holding key patents on genetic mutations used in TPMT testing, University of Manchester focusing on the issues related to the cost-effectiveness of using TPMT testing technology, while the Dr Margarete Fischer Bosch Institute — a large research foundation working on the customisation and improvement of drug therapy — hosted staff that held together a wider collaborative network through personal links (e.g. visiting fellowships) at other major centres in the field.¹³

The above discussed cases and techniques combined the geographical and social spaces of emergences and moved across units of analysis, in this case from the city- to the organisation-level. This provides a clear example of the flexibility and granularity of the overlay mapping in providing intelligence on the constellations of actors involved in the emergence process, structure of the relationships among these actors, key actors and collaborations shaping the emergence as well as main channels where knowledge and resources may flow.

4.3 Tracing emergence in the cognitive space

As new technologies emerge, epistemic developments occur in terms of discoveries, novel theories, or changes in technical developments such as experimental systems, materials, methods, and instrumentation (Joerges and Shinn, 2002; Rheinberger, 1997). These dynamics can be traced across the cognitive space by creating overlays of publications on basemaps of science that can be defined at different levels of analysis (e.g. Klavans and Boyack, 2009; Waltman and van Eck, 2012).

The publishing activity related to three case-studies can be, for example, projected across the map of science defined by the 225 WoS categories (e.g. Rafols et al., 2010). In this map, each node is a WoS category that can be assumed as proxy of a scientific discipline. The projection (overlay) makes a node’s size proportional to the number of publications related to the given

¹³The dynamic collaborative networks of the three case-studies are also included as supplementary materials.

technology that were published in the given discipline the node represents. The different colours of nodes represent different clusters of disciplines. Figure 6 depicts the projections of publications related to the three case-studies. We report the map representing the structure of science (left) — the strength of each linkage is proportional to the extent to which the two disciplines cite each other — and the heatmap version (right).¹⁴ This combination provides an intuitive visualisation of the diffusion process of emerging technologies.

As for the previous analyses, we used overlays projecting the publishing activity according to 5-year time windows. While these maps show the rapid diffusion of RNAi technology across many disciplines such as molecular biology, oncology, biomedical research, and chemistry, the overlays of HPV and TPMT testing technologies reveal different directions of diffusion. HPV testing technology diffuses from basic research in oncology, pathology, and virology disciplines towards issues related to the public health. We interpret this dynamic as a representation of the extensive and ongoing debate on the practices adopted for the screening of the population. The debate has been focused on the suitability of HPV testing technology as first an adjunct to, and latterly a substitute for, the widely adopted Pap test (Hogarth et al., 2012).

TPMP testing technology diffuses from the basic research in pharmacology towards, inter alia, gastroenterology and dermatology clinical disciplines. Publication activity seems to equally spread in gastroenterology and dermatology disciplines during the 1992-1996 period. Yet, in the subsequent years, the volume of publications shrinks from dermatology area while continuing to grow in gastroenterology. This raises questions about the degree to which TPMT's use has been contested in the different communities such as gastroenterologists, rheumatologists, and dermatologists.

It is worth noting that the techniques we have presented so far also suggest avenues of investigation an analyst may want to pursue for a more informed policy-making process. For example, we followed the aforementioned dynamic by conducting an additional analysis of the content of the scientific articles published by these communities. One potential conclusion may be that less publication in dermatology suggests failure to adopt. However, this is not the case. More fine-grained analysis of journal article titles and abstracts showed that while the community of dermatologists broadly accepted the validity of TPMT test, this test is still highly contested in the communities of rheumatologists and, especially, gastroenterologists. This more intense

¹⁴The visualisations of cognitive maps were produced by using VOSviewer 1.5.4 (van Eck and Waltman, 2010).

debate that is captured by the cognitive mapping approach is also testified by the strongly worded titles of scientific articles and presence of a large number of author response/reply or comment articles to other articles that these two communities published.

A similar cognitive perspective can be built by using a map of which nodes represent academic journals (Leydesdorff et al., 2013). The map is specifically composed by 10,330 journals (nodes) — the different colours of nodes represent different cluster of journals, i.e. groups of journals of which the cross-citation patterns are similar. Figure 7, for example, illustrates the rapid diffusion of RNAi across this map. RNAi has specifically started to appear in journals in the basic biomedical science and subsequently it has diffused among a variety of discipline-specific journals. The Rao-Stirling diversity index (Stirling, 2007), measured on the set of journals of the map, provides further evidence of this rapid diffusion, especially when the index is compared with the other two emerging technologies on which we focused our analysis.

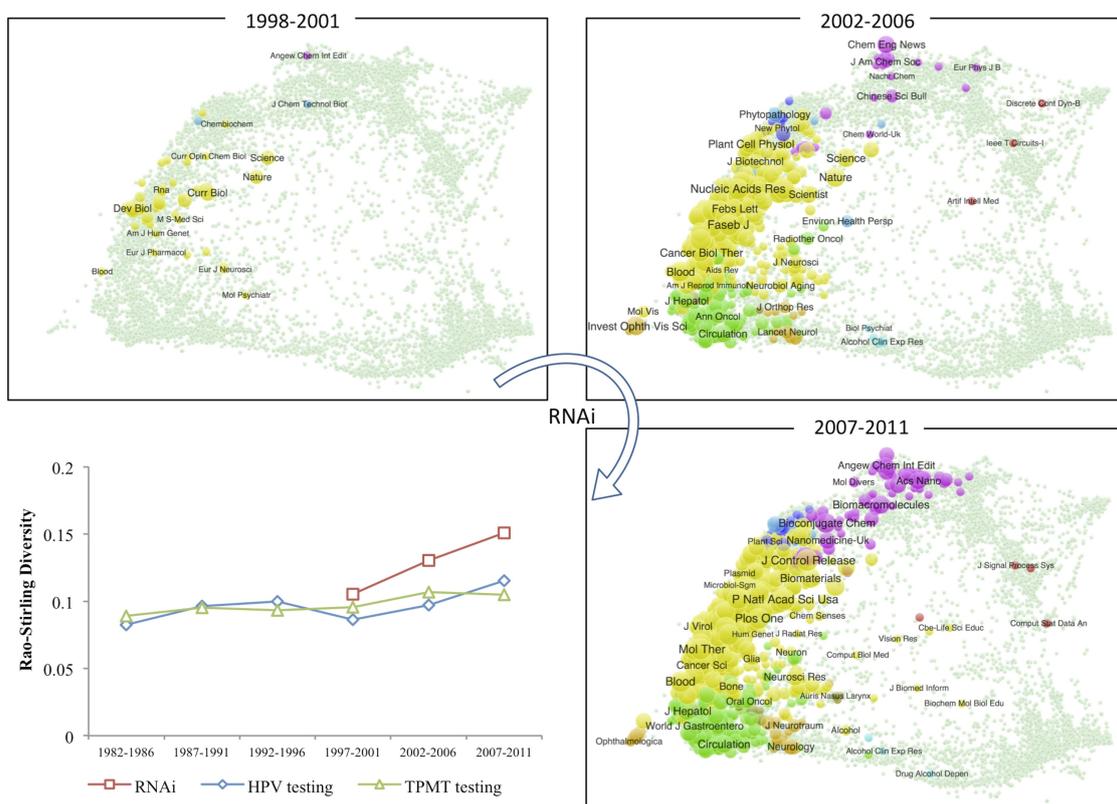


Figure 7: Journal map for RNAi and diversity index.

Perspectives on the cognitive dynamics can be also built by using MeSH terms — terms used to characterise the content of scientific articles in life science. These terms are assigned to articles in MEDLINE/PubMed through an intensive indexing process that is performed by examiners at the National Institute of Health (NIH). The terms are organised in a 16-branch

tree, which can reach up to 12 levels of depth. Drawing from this classification, Leydesdorff et al. (2012) developed a MeSH map by using three branches: "Diseases", "Chemicals and Drugs", and "Analytical, Diagnostics and Therapeutic Techniques and Equipment". These can be interpreted as 'supply-side', 'demand-side' and 'infrastructure', respectively. The first two levels of the tree were selected to build the basemap. The map is specifically composed by 822 MeSH terms (nodes) of which linkages reflect the (cosine) similarity according to co-occurrence of these terms in scientific articles. Each branch is marked on the map with a different colour: "Disease" is coloured red, "Chemicals and Drugs" green, and "Analytical, Diagnostics and Therapeutic Techniques and Equipment" blue (see Figure 8). Similarly to previous approaches, the publishing activity characterising a given emerging technology can be projected on this map to trace dynamics across three branches of the MeSH tree.

This approach, applied to the three case-studies, revealed different dynamics. RNAi, in line with previous results, 'globalises' across the set of the MeSH terms thus affecting many areas of the represented branches. On the contrary, HPV testing technology diffuses from the "Diseases" branch, specifically from "Tumor Virus Infections", into the "Analytical, Diagnostics and Therapeutic Techniques and Equipment" branch and eventually across the "Chemicals and Drugs" area. Yet, interestingly, in the last time window (2007-2011 period) scientific articles on HPV testing technology concentrate in the techniques and equipment area. This may reflect the efforts in developing competing HPV testing technologies. Results also show a specialisation of the TPMT case-study in specific areas of the map. This is in line with the scale of TPMT testing technology that is limited to a narrow domain of application.

Tracing the patenting activity of emerging technologies provides additional perspectives on the cognitive dynamics of the emergence process given the diverse incentives featuring in the creation process of scientific articles and patents. Scholars have developed techniques also to trace the dynamics of the patenting activities (e.g. Kay et al., 2014; Schoen et al., 2012). The nodes of these maps are technological classes that, as in the case of previous maps, are linked by cross-citation (cosine) similarity (Leydesdorff et al., 2014).

Figure 9, for example, depicts the overlays of RNAi patenting activity on the patent map based on technological areas as defined by the International Patent Classification (IPC). One can trace the dynamics in this space by moving across different levels of the classification (e.g. 3-digit, 4-digit). The patent map visualisation revealed the patenting activity of RNAi focused in specific areas of the technological space as biochemistry, organic chemistry, and medical science.

While one would expect to observe intense patenting in these technological areas given the nature of RNAi, this also reveals how one may need to increase, from case to case, the granularity of the mapping (i.e. lower levels of IPC classes) to build informative and interpretative perspectives on the observed emerging technology.

Mapping the emergence in the cognitive space may therefore reveal a number of relevant dynamics. These include the directions of diffusion of the given emerging technology across the key knowledge areas involved in emergence, how those areas may integrate or misalign, in which domains actors' knowledge production processes are positioned, or, for medical innovation, what diseases a technology is addressing and by using what type of techniques and chemicals.

5 Discussion

In cases of the emergence of novel technologies, all actors and even more so policy makers have very incomplete knowledge, not knowing, for example, where the boundaries of the technology are, in which direction it is moving, and how it should move forward. In the face of uncertainty, ambiguity, and ignorance (Stirling and Scoones, 2009), the emergence process should be investigated and analysed with strategic intelligence tools able to support a more informed policy-making process. In this article, we have presented a variety of overlay mapping approaches, applied those on three case-studies of emerging technologies in the biomedical domain, and shown the variety of 'intelligence' inputs for the policy-making process they can generate.

One of the advantages of those techniques is to not rely on single perspectives of the emergence, which would be misleading since relevant contexts would be inevitably omitted being emergence a complex and multi-faceted process. The mapping allows instead conduct a multi-perspective analysis on emergence and associated dynamics across the geographical, social, and cognitive spaces and combinations of these. Overlay mapping can be therefore conceived as a control system for emerging technologies that by monitoring emergence and synthesising the complex information deriving from it in relative simple visualisations can potentially favour the development of governance of a higher scope, speed, accuracy, and reliability (Nightingale, 2003). However, the flexibility and granularity of the mapping make its efficacy and efficiency to serve as a strategic intelligent tool dependent on a set of choices the analyst makes — some choices that may seem eminently technical (e.g. level of aggregation) may also have important implications for the type of patterns one can observe.

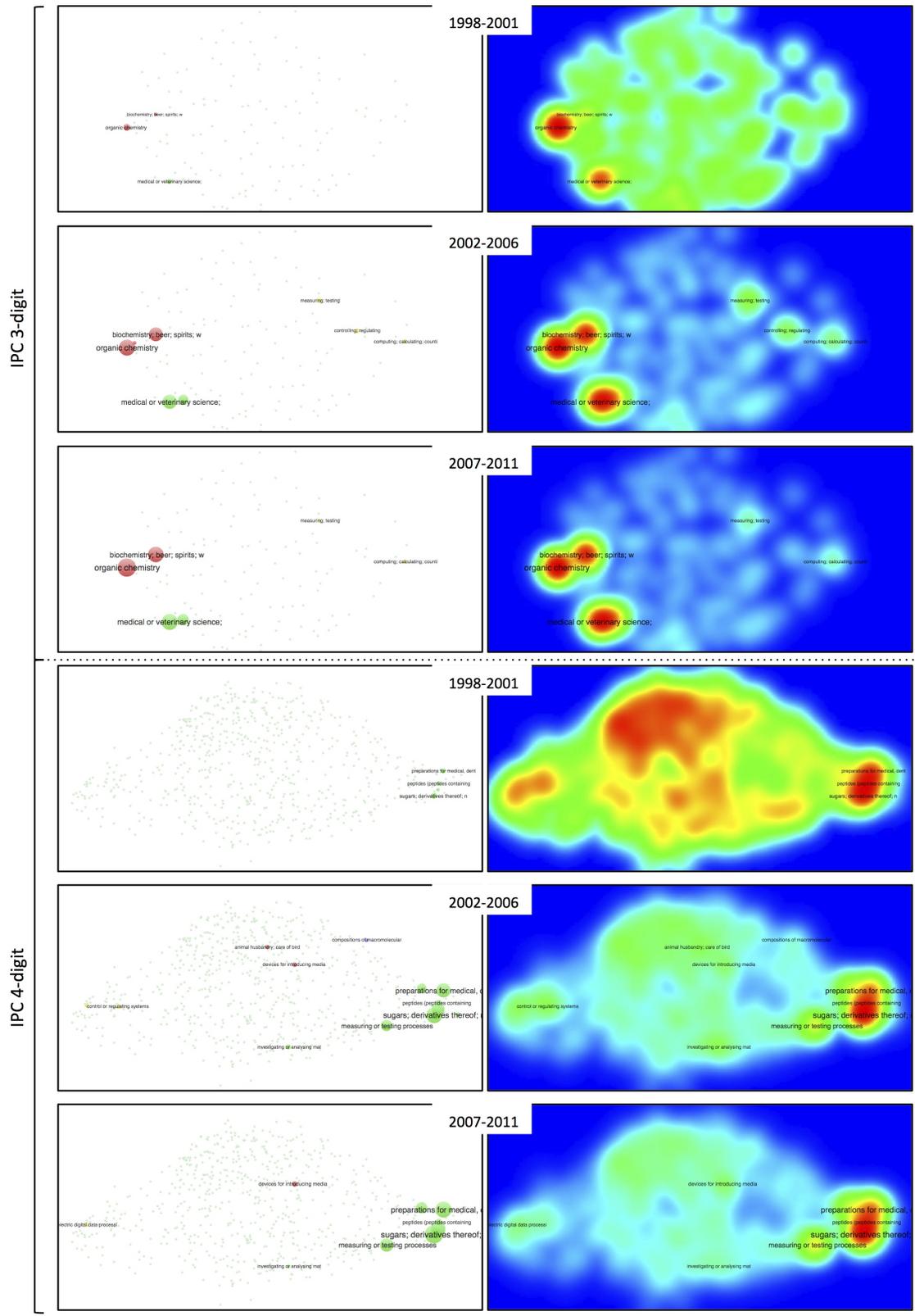


Figure 9: Patent map (IPC-based) for RNAi.

Firstly, a key preliminary choice is to identify the boundary of emerging technologies, i.e. the delineation of the corpus. Deciding what to include and what to exclude in the analysis is always a problematic exercise in the case of emergent sciences and technologies. In the case of RNAi, for example, one type of delineation portrayed this technology as already having reached a mature phase, whereas a broader definition suggested RNAi in a phase of rapid growth (Figure 2).

Secondly, the analyst has to identify the data sources. It is clear that using bibliographic data from publications and patents yields different types of information. Yet, those are only one form of research outputs. Many other possible dimensions such as products, services, changes in healthcare outcomes are not easily accessible for the mapping. This implies that some emerging technologies will be well represented in certain datasets and not others. This may also be the case for certain groups of actors — for example, academic organisations may publish more and patent less than private organisations, who may behave in the opposite manner. Consideration of available datasets places a considerable onus on the analyst to find appropriate sources and not to over-interpret limited data from less appropriate sources. Studying biotechnological innovations in an IP dominated industry, where regulatory and peer community pressures ensure publication, is a considerable advantage from the point of view of data access and availability. However, other contexts may not be so well provisioned, with implications for the utility of the approaches discussed here.

Thirdly, the selection of the elements to be analysed from the records of the databases as well as the categories into which the elements are assigned may have a significant impact on the resulting analyses. From a patent record, for example, one can extract information about inventors, firms, technology classes, or location, which, as discussed, provide insights on the social, cognitive or geographical spaces of emergence, respectively. In the case of categories one needs to decide the level of aggregation (the granularity of the description) as well as the type of classification — i.e., whether a predetermined, top-down typology is used (e.g., the MeSH terms) or an emerging, bottom-up taxonomy.

The choices of databases, elements, and categories can be informed by conceptual and theoretical frameworks that suggest why taking some perspective can be more fruitful than others to trace emerging technologies. However, from a policy analyst’s standpoint, it may be also critical to monitor technological emergence even when there is lack of an explicit understanding of conceptual frameworks used — policy is needed under conditions of incomplete knowledge. Yet, even in these cases in which there is no explicit adoption of a conceptual framework, the

choice of certain elements and categories in the analysis is privileging certain understanding over others, and this it is implicitly using the corresponding conceptual framework.

For example, looking at collaborative networks of individual scientists may place more attention on the social capital as a key factor in the emergence process (Nahapiet and Ghoshal, 1998). If one looks at the disciplinary position of the technology it is likely to be assuming that integration of disparate knowledge is relevant as, for example, it has been perceived to be the case of nanotechnologies (e.g. Porter and Rafols, 2009; Schummer, 2004) and in RNAi (Leydesdorff and Rafols, 2011), but not necessarily in HPV or TPMT testing technologies. In the absence of a clear understanding of the technology, one is advised to explore the phenomenon by using several perspectives, because one does not know in advance which one may turn out to be useful for understanding the relations of the emergence in place, or because the area in which the action is occurring is shifting over time.

Emerging technologies often do not conform to established bodies of knowledge. They cut across pre-existing organisational and institutional units, challenging established managerial and policy practices. As a result, one key demand from analysts and decision-makers is a description of the types of interdisciplinarity or convergence evident in emergence (Schmidt, 2007), often related with specific visions and expectations (Beckert et al., 2007; Roco and Bainbridge, 2002). In this regard, building multiple perspectives on the process of emergence requires moving across databases where emerging technologies 'tumble' in terms of representations with different attributes — such as geographical addresses, WoS Categories, patent classes.

The databases are retained and organized in different contexts with relative institutional rigidities. For example, relating patents to publications in terms of 'non-patent literature references' requires professional skills and cannot be done on a large scale without substantive investments. Thus, while the mapping and overlay techniques allow one using the same or a highly similar search strings across these databases more research is required in this direction to increase the integration of datasets. This can already be seen in attempts by US funders such as the NIH to link research grants, patents and publications.¹⁵

¹⁵See, for instance, <http://projectreporter.nih.gov/reporter.cfm>

6 Conclusions

We contributed to the debate of governance of emerging technologies by showing how overlay mapping can function as a tool of strategic intelligence for the governance of emerging technologies. The outputs produced by this stream of approaches can timely provide historical and informed perspectives on emergence with relative low efforts of analysts and decision-makers. Relevant dynamics are illustrated across the geographical, social, and cognitive spaces, in terms of geographical distribution of knowledge processes, collaborative interactions among geographical areas and organisational actors, and scientific and technological domains emerging or involved. Overlay mapping can also crosscut sources of data and move across units of analysis. These flexibility and granularity favour the comparison of the results from different policy-making contexts thus supporting the use of the mapping for a more 'distributed' strategic intelligence on emerging technologies.

We illustrated our argument conducting three case-studies of emerging technologies in the biomedical domains. The case-studies, being heterogeneous in terms of scale, applications, and phase of development, allowed us to explore the use of these techniques in distinct circumstances. Results showed how the overlay mapping approach provides both overview and the possibility to focus by zooming in on specific developments and phenomena. For example, in the case of HPV and TPMT testing technologies, we showed how different facets and dynamics of the emergence process can be revealed by moving across the geographical and social spaces from the city-level to organisation-level units of analysis. The maps also enable the user to compare different technologies and alternatives in terms of strengths and weaknesses at the portfolio level. As in the case of RNAi, the mapping can reveal cases of emerging technologies initially concentrated at a few places with major players hosting the sets of relevant capabilities, competences, and network relationships, and then diffusing across a number of actors according to preferential attachment mechanisms (Leydesdorff and Rafols, 2011). On this 'journey' of the emerging technology, changes in the involved domains of science and technology — especially in the case of RNAi, the diffusion of laboratory research tools in other specialties and disciplines — can also be revealed with the cognitive mapping.

Notwithstanding the delineation issues associated with emerging technologies, the need of some qualitative background on the studied technology for the interpretation and refinement of the mapping, the limitations of the data that can be used (mainly publications and patents),

overlay mapping can function as heuristic that allow confronting the respective theoretical debates with puzzles in the relevant data as well as informing governance by specifying uncertainties in considerable detail. The mapping may also pose additional questions to the analyst and decision-maker. The produced perspectives on emergence, according to the set of choices an analyst takes, may timely suggest directions of further investigation as well as feeding into political discourse about stumbling blocks ahead and possible openings in the landscapes.

Acknowledgements

The authors acknowledge the support of UK Economic and Social Research Council (award RES-360-25-0076 - "[Mapping the Dynamics of Emergent Technologies](#)") in the development of the case-studies. The authors are also grateful to the US National Science Foundation (award #1064146 - "[Revealing Innovation Pathways: Hybrid Science Maps for Technology Assessment and Foresight](#)") for the support in the development of the mapping techniques. The findings and observations contained in this paper are those of the authors and do not necessarily reflect the funders' views. The authors are grateful to Paul Nightingale, Stefan Kuhlmann, and Peter Stegmaier for their feedbacks on earlier versions of this paper. Previous versions of this paper were presented at the 2013 ISSI Conference, 2013 EU-SPRI conference, 2013 Atlanta Conference, 2013 DRUID Conference, and "2013 Data Science, Tech Mining and Innovation" workshop in Manchester.

References

- Beckert, B., Blümel, C., and Friedewald, M. (2007). Visions and realities in converging technologies. Exploring the technology base for convergence. *Innovation: The European Journal of Social Science Research*, 20(4):375–194.
- Börner, K. (2010). *Atlas of Science. Vizualizing What We Know*. MIT Press, Cambridge, MA.
- Bornmann, L. and Leydesdorff, L. (2011). Which cities produce more excellent papers than can be expected? A new mapping approach, using Google Maps, based on statistical significance testing. *Journal of the American Society for Information Science and Technology*, 62(10):1954–1962.

- Bornmann, L. and Waltman, L. (2011). The detection of hot regions in the geography of science: A visualization approach by using density maps. *Journal of Informetrics*, 5(4):547–553.
- Callon, M., Courtial, J. J. P., Turner, W. A., and Bauin, S. (1983). From translations to problematic networks: An introduction to coword analysis. *Social Science Information*, 22(2):191–235.
- Cambrosio, A., Keating, P., Mercier, S., Lewison, G., and Mogoutov, A. (2006). Mapping the emergence and development of translational cancer research. *European Journal of Cancer*, 42(18):3140–3148.
- Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, 57(3):359–377.
- Collingridge, D. (1980). *The Social Control of Technology*. Open University Press, Milton Keynes.
- Corrocher, N., Malerba, F., and Montobbio, F. (2003). The emergence of new technologies in the ICT field: main actors, geographical distribution and knowledge sources. *Working Papers of Faculty of Economics, Universita degli Studi dell’Insubria*.
- Cozzens, S. E., Gatchair, S., Kang, J., Kim, K.-S., Lee, H. J., Ordóñez, G., and Porter, A. (2010). Emerging technologies: quantitative identification and measurement. *Technology Analysis & Strategic Management*, 22(3):361–376.
- Crane, D. (1972). *Invisible Colleges: Diffusion of Knowledge in Scientific Communities*. The University of Chicago Press, Chicago.
- David, P. (1985). Clio and the economics of QWERTY. *American Economic Review*, 75(2):332–337.
- Day, G. S. and Schoemaker, P. J. H. (2000). Avoiding the pitfalls of emerging technologies. *California Management Review*, 42(2):8–33.
- de Haan, J. (2006). How emergence arises. *Ecological Complexity*, 3(4):293–301.
- Fire, A., Xu, S., Montgomery, M. K., Kostas, S. A., Driver, S. E., and Mello, C. C. (1998). Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. *Nature*, 391(6669):806–811.

- Garfield, E. (1979). *Citation Indexing: Its Theory and Application in Science, Technology, and Humanities*. John Wiley, New York.
- Garfield, E. (1987). Launching the ISI Atlas of Science: for the new year, a new generation of reviews. *Current contents*, 1:1–6.
- Geels, F. W. (2002). Technological transitions as evolutionary reconfiguration processes: a multi-level perspective and a case-study. *Research Policy*, 31(8-9):1257–1274.
- Giddens, A. (1984). *The Constitution of Society*. University of California Press, Berkeley, CA.
- Glänzel, W. and Schubert, A. (2004). Analysing scientific networks through co-authorship. In Moed, H., Glänzel, W., and Schmoch, U., editors, *Handbook of Quantitative Science and Technology Research*, pages 257–276. Kluwer Academic Publishers, Dordrecht.
- Griliches, Z. (1994). Productivity, R&D, and the data constraint. *American Economic Review*, 84(1):1–23.
- Guston, D. D. H. and Sarewitz, D. (2002). Real-time technology assessment. *Technology in Society*, 24(1-2):93–190.
- Haussecker, D. (2008). The business of RNAi therapeutics. *Human Gene Therapy*, 19(5):451–462.
- Haussecker, D. (2012). The business of RNAi therapeutics in 2012. *Molecular Therapy Nucleic Acids*, 1(2):e8.
- Hilgartner, S. and Lewenstein, B. (2014). The speculative world of emerging technologies. *Working paper of Cornell University*.
- Hogarth, S., Hopkins, M. M., and Rodriguez, V. (2012). A molecular monopoly? HPV testing, the Pap smear and the molecularisation of cervical cancer screening in the USA. *Sociology of Health & Illness*, 34(2):234–250.
- Hopkins, M. M., Ibarreta, D., Gaisser, S., Enzing, C. M., Ryan, J., Martin, P. A., Lewis, G., Detmar, S., Marle, M. E. v. d. A.-v., Hedgecoe, A. M., Nightingale, P., Dreiling, M., Hartig, K. J., Vullings, W., and Forde, T. (2006). Putting pharmacogenetics into practice. *Nature Biotechnology*, 24(4):403–410.

- Hu, G., Carley, S., and Tang, L. (2012). Visualizing nanotechnology research in Canada: evidence from publication activities, 1990-2009. *Journal of Technology Transfer*, 37(4):550–562.
- Hunt, D., Nguyen, L., and Rodgers, M., editors (2007). *Patent Searching: Tools and Techniques*. John Wiley & Sons, Hoboken, New Jersey.
- Irvine, J. and Martin, B. R. (1985). Basic research in the East and West: A comparison of the scientific performance of high-energy physics accelerators. *Social Studies of Science*, 15(2):293–341.
- Jaffe, A. B. and Trajtenberg, M. (2002). *Patents, Citations & Innovations: A Window on the Knowledge Economy*. MIT Press, Cambridge, MA.
- Joerges, B. and Shinn, T. (2002). *Instrumentation Between Science, State and Industry*. Kluwer Academic Publishers, Dordrecht.
- Kay, L., Newman, N., Youtie, J., Porter, A. L., and Rafols, I. (2014). Patent overlay mapping: Visualizing technological distance. *Journal American Society for Information Science and Technology*, in press.
- Kessler, M. (1963). Bibliographic coupling between scientific papers. *American Documentation*, 14(1):10–25.
- Klavans, R. and Boyack, K. W. (2009). Toward a consensus map of science. *Journal of the American Society for Information Science and Technology*, 60(3):455–476.
- Kooiman, J. (1993). *Modern Governance. New Government-Society Interactions*. Sage, London.
- Kuhlmann, S., Boekholt, P., Georghiou, L., Guy, K., Heraud, J.-A., Laredo, P., Lemola, T., Loveridge, D., Luukkonen, T., Moniz, A., Polt, W., Rip, A., Sanz-Menendez, L., and Smits, R. (1999). Improving distributed intelligence in complex innovation systems. *MPRA Working Paper*.
- Latour, B. (1993). *The Pasteurization of France*. Harvard University Press, Cambridge, MA.
- Leydesdorff, L. (1989). Words and co-words as indicators of intellectual organization. *Research Policy*, 18(4):209–223.

- Leydesdorff, L. and Bornmann, L. (2012). Mapping (USPTO) patent data using overlays to Google Maps. *Journal of the American Society for Information Science and Technology*, 63(7):1442–1458.
- Leydesdorff, L., Kushnir, D., and Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on International Patent Classification (IPC). *Scientometrics*, 98(3):1583–1599.
- Leydesdorff, L. and Persson, O. (2010). Mapping the geography of science: Distribution patterns and networks of relations among cities and institutes. *Journal of the American Society for Information Science and Technology*, 61(8):1622–1634.
- Leydesdorff, L. and Rafols, I. (2011). The Local Emergence and Global Diffusion of Research Technologies: An Exploration of Patterns of Network Formation. *Journal of the American Society for Information Science and Technology*, 62(5):846–860.
- Leydesdorff, L., Rafols, I., and Chaomei, C. (2013). Interactive overlays of journals and the measurement of interdisciplinarity on the basis of aggregated journal-journal citations. *Journal American Society for Information Science and Technology*, 64(12):2573–2586.
- Leydesdorff, L., Rotolo, D., and Rafols, I. (2012). Bibliometric perspectives on medical innovation using the medical subject Headings of PubMed. *Journal of the American Society for Information Science and Technology*, 63(11):2239–2253.
- Lundin, P. (2011). Is silence still golden? Mapping the RNAi patent landscape. *Nature Biotechnology*, 29(6):493–497.
- Martin, B. R. (1995). Foresight in science and technology. *Technology Analysis & Strategic Management*, 7(2):139–168.
- Murray, F. (2002). Innovation as co-evolution of scientific and technological networks: exploring tissue engineering. *Research Policy*, 31(8-9):1389–1403.
- Nahapiet, J. and Ghoshal, S. (1998). Social capital, intellectual capital, and the organizational advantage. *Academy of Management Review*, 23(2):242–266.
- Narin, F. (1976). *Evaluative Bibliometrics: The Use of Publication and Citation Analysis in the Evaluation of Scientific Activity*. National Science Foundation, Washington, DC.

- Nightingale, P. (2003). Capacity utilization revisited: software, control and the growth of large technical systems. *Industrial and Corporate Change*, 12(3):477–517.
- Noyons, C. (2005). Science maps within a science policy context. In Moed, H., Glänzel, W., and Schmoch, U., editors, *Handbook of Quantitative Science and Technology Research*, pages 237–255. Kluwer Academic Publishers, Dordrecht.
- Noyons, E. (2001). Bibliometric mapping of science in a science policy context. *Scientometrics*, 50(1):83–98.
- Ohniwa, R. L., Hibino, A., and Takeyasu, K. (2010). Trends in research foci in life science fields over the last 30years monitored by emerging topics. *Scientometrics*, 85(1):111–127.
- Porter, A. and Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, 81(3):719–745.
- Powell, W. W. (1990). The Transformation of Organizational Forms: How Useful is Organisation Thoery in Accounting for Social Change. In Friedland, R. O. and Robertson, A., editors, *Beyond the Marketplace: Rethinking Economy and Society*. Walter de Gruyter, New York, NY, USA.
- Price, D. (1965). Is technology historically independent of science? A study in statistical historiography. *Technology and Culture*, 6(4):553–568.
- Rafols, I., Porter, A. L., and Leydesdorff, L. (2010). Science overlay maps: A new tool for research policy and library management. *Journal of the American Society for Information Science and Technology*, 61(9):1871–1887.
- Rheinberger, H.-J. (1997). *Toward a History of Epistemic Things: Synthesizing Proteins in the Test Tube*. Stanford University Press, Palo Alto, CA.
- Rip, A. and Van Amerom, M. (2010). Emerging De Facto Agendas Surrounding Nanotechnology: Two Cases Full of Contingencies, Lock-outs, and Lock-ins. In Kaiser, M., Kurath, M., Maasen, S., and Rehmann-Sutter, C., editors, *Governing Future Technologies*, volume 27 of *Sociology of the Sciences Yearbook*, pages 131–155. Springer Netherlands, Dordrecht.
- Roco, M. and Bainbridge, W. (2002). *Converging technologies for improving human performance. Nanotechnology, biotechnology, information technology and cognitive science*. NSF, Arlington, VA, US.

- Schmidt, J. (2007). Knowledge politics of interdisciplinarity. Specifying the type of interdisciplinarity in the NSF's NBIC scenario. *Innovation: The European Journal of Social Science Research*, 20(4):313–328.
- Schoen, A., Villard, L., Laurens, P., Cointet, J.-P., Heimeriks, G., and Alkemade, F. (2012). The network structure of technological developments; Technological distance as a walk on the technology map. In *2012 Science & Technology Indicators (STI) Conference*, Montreal (Canada).
- Schummer, J. (2004). Multidisciplinarity, interdisciplinarity, and patterns of research collaboration in nanoscience and nanotechnology. *Scientometrics*, 59(3):425–465.
- Small, H. (1973). Co-citation in the scientific literature: a new measure of the relationship between two documents. *Journal of the American Society for Information Science*, 24(4):265–269.
- Small, H., Boyack, K. W., and Klavans, R. (2014). Identifying emerging topics in science and technology. *Research Policy*, in press.
- Small, H. and Garfield, E. (1985). The geography of science: disciplinary and national mappings. *Journal of Information Science*, 11(4):147–159.
- Stirling, A. (2007). A General framework for analysing diversity in science, technology and society. *Journal of The Royal Society Interface*, 4(15):707–719.
- Stirling, A. C. and Scoones, I. (2009). From risk assessment to knowledge mapping: Science, precaution, and participation in disease ecology. *Ecology and Society*, 14(2).
- Sung, J. J. and Hopkins, M. M. (2006). Towards a method for evaluating technological expectations: Revealing uncertainty in gene silencing technology discourse. *Technology Analysis & Strategic Management*, 18(3-4):345–359.
- van Eck, N. J. and Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2):523–538.
- Van Lente, H. and Rip, A. (1998). The rise of membrane technology. From rethorics to social reality. *Social Studies of Science*, 28(2):221–254.

- Verbong, G. and Geels, F. (2007). The ongoing energy transition: Lessons from a socio-technical, multi-level analysis of the Dutch electricity system (1960-2004). *Energy Policy*, 35(2):1025–1037.
- Wagner, C. S. (2008). *The New Invisible College*. Brookings Institution Press, Washington, DC.
- Waltman, L. and van Eck, N. J. (2012). A new methodology for constructing a publication-level classification system of science. *Journal American Society for Information Science and Technology*, 63(12):2378–2392.
- Wassermann, S. and Faust, K. (1994). *Social Network Analysis: Methods and Applications*. Cambridge University Press, Cambridge, MA.
- Zitt, M. and Bassecoulard, E. (2006). Delineating complex scientific fields by an hybrid lexical-citation method: An application to nanosciences. *Information Processing & Management*, 42(6):1513–1531.
- zur Hausen, H. (1987). Papillomaviruses in human cancer. *Applied Pathology*, 5:19–24.

Recent papers in the SPRU Working Paper Series:

SWPS 2013-16. Mariana Mazzucato and Stuart Parris. November 2013. "High growth firms, innovation and competition: the case of the US pharmaceutical industry".

SWPS 2013-17. Hakim Hammadou, Sonia Paty, and Maria Savona. December 2013. "Strategic interactions in public R&D across European countries: A spatial econometric analysis".

SWPS 2014-01. Nicola Grassano and Maria Savona. January 2014. "Productivity in services twenty years on. A review of conceptual and measurement issues and a way forward".

SWPS 2014-02. Florian Kern, Adrian Smith, Chris Shaw, Rob Raven and Bram Verhees. February 2014. "From laggard to leader: Explaining offshore wind developments in the UK."

SWPS 2014-03. Andy Stirling . March 2014. "Transforming Power: social science and the politics of energy choices."

SWPS 2014-04. Alex Coad, Gabriele Pellegrino and Maria Savona. March 2014. "Don't Stop Me Now: Barriers to innovation and firm productivity."

SWPS 2014-05. Antonio Messeni Petruzzelli, Vito Albino, and Daniele Rotolo. April 2014. "Determinants of Patent Citations in Biotechnology: An Analysis of Patent Influence Across the Industrial and Organizational Boundaries."

SWPS 2014-06. Andy Stirling. April 2014. "From Sustainability to Transformation: Dynamics and diversity in reflexive governance of vulnerability."

SWPS 2014-07. Ralitsa Petrova Hiteva and Tomas Maltby. April 2014. "Standing in the way by standing in the middle: the case of state-owned natural gas intermediaries in Bulgaria."

SWPS 2014-08. Sabine Hielscher and Adrian Smith. May 2014. "Community-based digital fabrication workshops: A review of the research literature."

SWPS 2014-09. José García-Quevedo, Gabriele Pellegrino and Maria Savona. May 2014. "Reviving demand-pull perspectives: The effect of demand uncertainty and stagnancy on R&D strategy."