Early-stage detection of breakthrough-class scientific research : using micro-level citation dynamics

Winnink, J.J.

Author: Winnink, J.J.
Title: Early-stage detection of breakthrough-class scientific research : using micro-level citation dynamics
Issue Date: 2017-02-22

Chapter 2

Case study #1: Integrase Inhibitors

The content of this chapter is equivalent to the manuscript published as:

Winnink, J.J. and Tijssen, R.J. (2014). R&D dynamics and scientific breakthroughs in HIV/aids drugs development: the case of integrase inhibitors. *Scientometrics*, 101(1):1–16. (DOI:10.1007/s11192-014-1330-7).

The only differences between the published article and the text in this chapter are of typographic nature to assure that the layout is in line with the typographic design of this PhD thesis.

Abstract

Transformations and applications of scientific knowledge into new technologies are usually complex interactive processes. Is it possible to detect from bibliographic information alone, structural alterations and significant events within these processes that may indicate breakthrough discoveries? In this empirical study we focus on R&D processes leading to HIV/AIDS medicines called Integrase Inhibitors. Where scientific progress and discoveries are reflected in research papers, patents signify inventions and technological achievements. Our temporal analysis of distinctive events in this R&D area, tracing trends within both bibliographic information sources, is driven by three bibliometric indicators: (1) contributions of 'bridging researchers' who are also inventors, (2) 'key papers' that subject experts in the field considered milestones in the research process, and (3) multidisciplinary impact of those papers. The main results indicate that a combination of key papers, bridging researchers and multidisciplinary impact might help track potential 'Charge type' breakthrough developments.

2.1 Introduction

This paper reports on the development of a statistical method for data-mining bibliographic information in order to describe and analyse 'distinctive' events or transformative developmental patterns within R&D processes. The main goal is to assess whether or not such events or transformations represent 'breakthroughs', i.e. the introduction of remarkably new knowledge or radically innovative technologies. There is growing academic literature discussing the characteristics of such breakthroughs, e.g. (Ahuja and Lampert, 2001; Chandy and Tellis, 2000; Chen et al., 2009; Henderson and Clark, 1990; Hollingsworth, 2008; Podolny and Stuart, 1995; Rosenkopf and Nerkar, 2001; Trajtenberg, 1990a,b; Tushman and Anderson, 1986). In this empirical study we focus our attention on scientific discoveries that are the cause of, or significantly contribute to, subsequent technological developments.

Scientists, researchers and scholars are often able, retrospectively, to identify documented evidence of such discoveries or inventions that were of crucial importance in these R&D processes. Such a document will be referred to as a *'key publication'*. These particular publications provide empirical information to help trace and identify transformative stages within R&D processes where a breakthrough occurred, and help describe characteristic features and structural determinants of processes that played a crucial role. In this study we examine the development of HIV/AIDS medicines, more specifically the research area of 'Integrase Inhibitors'. Our main research question is: 'Which indicator or indicators, based on bibliographic data of scholarly publications and patent publications, could be applied in early stage identification of potential scientific breakthroughs?' Section 2.2 'Analytical framework' first briefly introduces previous work on identifying breakthroughs, discusses the concept of 'bridging researchers', provides a short overview of the development of HIV/AIDS medicines, and argues the decision to focus in this study on the development of Integrase Inhibitor medicines. We further describe our bibliographic information sources, and discuss how longitudinal developments were analysed. The main results in Section 2.3 'Results' highlight highly cited key publications in this research area, and the multidisciplinary impact against the backdrop of the most prominent bridging researchers. Section 2.4 'Discussion and concluding remarks' summarizes and concludes with general observations.

2.2 Analytical framework

2.2.1 Defining and identifying breakthroughs

There are many conceptual and methodological challenges involved in identifying scientific breakthroughs, and pinning down the moment within the R&D process when they occurred. Various detection methods and analytical tools already have been tried and tested to solve this problem — with limited success (Arbesman, 2010; Bettencourt et al., 2008; Breiner et al., 1994; Julius et al., 1977; Martin, 1995). More recently, Chen (2012) takes an 'informational' perspective by focussing on the question 'What kinds of information may serve as early signs of potentially valuable ideas?'; Small and Klavant (2011) combine co-citation analysis and the context of citations; Ponomarev et al. (2012) provides an illustrative example of this data-informed and evidence-based approach by means of a two-step citation impact methodology applied to research publications.[44]

Our informational study also operationalizes (potential) breakthroughs in terms of bibliographic data, but from a broader range, which includes research publications and patents. In our longitudinal study time is of the essence. We define an event time-line based on the publication date of the research publication and the oldest priority date of patents.[45] Within this framework we characterize key events by the emergence of one or more key papers that were produced by bridging researchers. These papers are either very highly cited, and/or exhibit a broad multidisciplinary citation impact, within 12 months after publication. Our underlying conceptual framework is based on Hollingsworth's description of a breakthrough 'A major breakthrough or discovery is a finding or process, often preceded by numerous small advances, which leads to a new way of thinking about a problem' (Hollingsworth, 2008), and on Koshland's Cha–Cha–Cha typology of scientific discoveries (Koshland, 2007).[46] Koshland distinguishes three types of discoveries: '**Cha**llenge' discoveries are a response to an accumulation of facts or concepts that remained unexplained by, or were incongruous with, scientific theories of the time, 'paradigm shifts' (Kuhn, 1962); 'Chance' discoveries suddenly appear and are often referred to as serendipitous, (van Andel, 1994) denotes discoveries of this type 'unsought findings'; 'Charge' discoveries solving known problems within an existing theoretical framework can be considered, in Kuhnian sense, evolutionary 'normal' science. Making a distinction between charge discoveries that do not involve a paradigm shift and challenge and chance discoveries that do is relevant as paradigm-changing discoveries have notoriously limited early impact (Redner, 2005), and are therefore hard to detect on the basis of early citation patterns (Wang et al., 2013). The Integrase Inhibitors breakthrough is a *Charge* discovery; the biochemical mechanism was already identified and no new theoretical understanding was needed — finding the right chemical compound to do the job was the main problem.

2.2.2 Bridging researchers

In addition to citation-based linkages between key publications, one can also connect documents through authorship. The same people can act as (co)

[44] Tracking the citation patterns over a period of 6–24 months after publication, and using this information to predict if a publication exceeds a citation threshold level within 5 years after publication.

[45] Oldest priority data is the date of the earliest filing of a patent application for this invention.

[46] Ciechanover (2009) applies Koshland's criteria to motivate the classification of the discovery of 'ubiquitin proteolytic system' as a '**Cha**llenge' rather than a '**Cha**nce'.

author of a research publication and as (co-)inventor of a patent. These author/inventors constitute an explicit link ('bridge') between knowledge creation and knowledge application. We argue that these 'bridging researchers' may play a pivotal role in overcoming barriers between basic research, applied research, and technological development in the development of radically new medicines. They are referred to as 'core scientists' (Furukawa and Goto, 2006) or 'Pasteur scientists' (Stokes, 1997) owing to their ability to recognize and reconcile different stages in the R&D chain. They have been subject of several studies of knowledge transfer mechanisms, where these individuals are considered to be an important thrust in moving scientific breakthroughs towards science-based innovations (Balconi et al., 2004; Lissoni, 2010; Meyer, 2006; Noyons et al., 1994; Tijssen, 2002; Zucker and Darby, 1996).

2.2.3 Development of new HIV/IDS medicines

Usually many years may pass before scientific discoveries may find their way into technological applications (Isenson, 1969; Jewkes et al., 1958). In medicine, it often takes 10–14 years from 'upstream' disease characterization to a 'downstream' drug approved for human usage (IOM, 2009). The costs involved in developing a new medicine from start to approval are estimated (FDA, 2004) to be in the range from US$ 0.8 to 1.7 billion. The public sector is key contributor in the process of drug development in the sense that universities or government laboratories conducts 'enabling research' (Cockburn and Henderson, 1998). In the 1980s a novel method for drug design known as 'Rational Drug Design' (RDD)[47] replaced then existing trial-and-error 'harvesting' methods for finding new 'lead compounds'[48] that could act as a basis for a new medicine. RDD's encouraged the development of various computational methods using structural information suggesting novel molecular structures. Each of these structures may either prove to be useful as a 'lead compound' or act as a stimulus to the creativity of designers (Wang et al., 2000). Using RDD the route from scientific knowledge to applicable HIV/AIDS medicines was traversed in a relatively short period of time and the development of these medicines has been very dynamic since the very start in the early 1980s.

HIV/AIDS medicines are currently classified into five[49] distinct classes based on the bio-molecular mechanisms used to attack the pathogen. In his review paper De Clercq (2009a) identifies a series of key publications describing developments considered being pivotal for the developments in the classes of HIV/AIDS medicines. Owing to the controlled and sequential nature of RDD-based development processes of new medicines, we should be able to identify and study scientific breakthrough 'transformative' events. We focus on R&D for medicine in the class of Integrase Inhibitors and specific on a

[47] Also known as 'computer-aided drug design', or 'structure-based drug design',

[48] Lead compounds are, if proven to be an effective drug, the beginning of a group of chemical closely related medicines

[49] A sixth class is formed by the so called cocktail drugs; combinations of two or more drugs from the other five classes.

medicine with the generic name Raltegravir, which is also known by its brand name *Isentress*®. It was the only approved drug of this class at the time of this study.

2.2.4 Focus on Integrase Inhibitors

The rationale to study the development of Integrase Inhibitors medicine is based on the following considerations: (1) the key paper that is the basis of the *Integrase Inhibitors* medicines (Hazuda et al., 2000) stands out, as it is the only 'citation classic' among the HIV/AIDS key papers (see Section 2.3), (2) at the time of this study *Isentress*® was the only approved drug of this class, and (3) the relative short period between the publication of the key paper on January 28, 2000 and the publication of the granted patent on January 30, 2007. Especially (2) and (3) restricted the number of scholarly and patent publications in the analysis, and therefore reducing the complexity of the study.

2.2.5 Bibliographic data

We selected bibliographical data for relevant scholarly publications from the period 1980–2010 from our Thomson Reuters/CWTS Web of Science database (TR/CWTS WoS). This dataset contains bibliographical information on 34,665 scholarly publications consisting of articles and letters related to the disease HIV/AIDS and to the development of HIV/AIDS medicine. We restricted the dataset to a number of subject categories[50] in order to exclude references to non-pharmaceutical or non-medical subject areas.

The patent data were extracted from the April 2012 release of the EPO PATSTAT database[51] covering the period 1980–2008 consisting of 16,379 distinct so called DOCDB patent families[52]. Each patent family was counted only once. We constructed the set of patent publications by applying patent classification codes related to HIV/AIDS, which was supplemented by with documents containing the terms 'HIV', or 'AIDS'. We imposed the constraint that the patent publications were also classified as 'medical preparation' (i.e. a medicine). In this set of documents we identified all patents belonging to the *Integrase Inhibitors* class.

Patent publications frequently contain references to non-patent publications. Part of those non-patent references (NPR's) consists of references to scholarly publications, which reflect direct or indirect links between the cited scientific knowledge and the citing patented technology, see e.g.(Grupp, 1992;

[50] Immunology, biochemistry & molecular biology, virology; cell biology, microbiology, infectious diseases, pharmacology & pharmacy, biotechnology & applied microbiology, medicinal chemistry, infectious diseases, and medicine

[51] The PATSTAT database is the database also known as the *EPO Worldwide Patent Statistical Database*.

[52] A DOCDB patent family is a collection of closely related equivalent patent publications describing the same invention. 2008 was used as cut-off year because due to procedural regulations more recent patent data were not yet fully available.

Hicks, 2000; Narin and Olivastro, 1998; Tijssen, 2002). These NPRs were used to study the temporal dynamics of science being utilized and transformed into patented applications and technological breakthroughs.

In order to track micro-level developments over time we need to synchronize the time line between bibliographic evidence. For patents we adopted 'the oldest priority date', the available date that is closest to the moment of the invention. For scholarly publications we selected the date a manuscript was received by the publisher, acknowledging the fact that this date is a crude proxy for the moment the breakthrough idea actually came about.[53]

Our case study focuses on one key publication that played a pivotal role in this development: the 'Hazuda 2000' paper that published the solution to the then remaining scientific problems (Hazuda et al., 2000). We identified all patent publications that contain a literature reference to this publication, which enabled us to compile a set of patent documents directly influenced by this publication. Starting with the patent publication on *Isentress*® (US patent 7169780) we created a backward citing-cited network and collected the publication years for the referenced publications. In the Web of Science, we selected publications published from 1999 onwards, which allowed us to include information prior to the hard-copy publication of 'Hazuda 2000'. Drawing on these sources, all inventors, authors, patent applicants, companies or research institutions active in the *Isentress*® R&D area were identified and incorporated in a longitudinal network structure.

2.3 Results

General

HIV was identified 1983 as the retrovirus causing the disease AIDS. The first research papers related to the development of HIV/AIDS medicine were published in 1981. Following the initial HIV/AIDS medicines five new classes of medicines based on new insights in the bio-molecular workings were developed. For each of these classes De Clercq's authoritative, Table 2.3 contains a short biography, review highlights the research publications that mark pivotal developments (De Clercq, 2009b). Table 2.1 lists all seven of De Clercq's key publications which we used as a frame of reference to assess (Hazuda et al., 2000) as a possible breakthrough. Three publications (Baba et al., 1989; Miyasaka et al., 1989; Pauwels et al., 1990) constitute the start of the non-nucleoside reverse transcriptase inhibitors.

Figure 2.1 shows the number of citations received by the key papers and an (weighted) average HIV/AIDS publication received the first 10 years after

[53] This manuscript submission date is however not provided in the databases. For individual publications the submission date of the manuscript can usually be found on the 'image' of the publication, but for analysis of larger datasets the use of this date is very laborious and time-consuming, and we therefore had to use the 'publication date' provided by the Web of Science. Further research has to be done in this area to estimate the error introduced; (Luwel and van Wijk, 2012) provides some insights into the nature and extent of this problem.

Table 2.1: General description used by De Clercq (2009a) to mark pivotal developments supplemented by the number of citations received by the key papers

Publication	Description of the pivotal development (De Clercq, 2009a)	Citations received		
		From scholarly publications		From patent publications
		During first 12 months	Total (ultimo 2010)	Total (ultimo 2008)
Baba et al. (1989)[a]	Discovery of chemical substance later to be known as HEPT; start of so called non-nucleoside reverse transcriptase inhibitors (NNRTIs)	3	265	0
Miyasaka et al. (1989)	Discovery of chemical substance later to be known as HEPT; start of so called non-nucleoside reverse transcriptase inhibitors (NNRTIs)	7	201	8
Pauwels et al. (1990)	Discovery of chemical substance later known as TIBO; start of so called non-nucleoside reverse transcriptase inhibitors (NNRTIs)	29	571	26
Roberts et al. (1990)	Discovery of Saquinavir, first of the proteinase inhibitors (PIs)	32	570	88
Wild et al. (1993)	Discovery of Enfuvirtide, first drug in the class fusion inhibitors (FIs)	3	115	12
Hazuda et al. (2000)	Scientific base for integrase inhibitors (INIs)	16	585	9
Esté and Telenti (2007)[b]	Discovery of mechanism on which co-receptor inhibitors (CRIs) are based	22	114	0

[a] Baba et al. (1989); Miyasaka et al. (1989) both concern the same finding (HEPT)

[b] Esté and Telenti (2007) is actually a review article, not specifically dealing with a certain drug

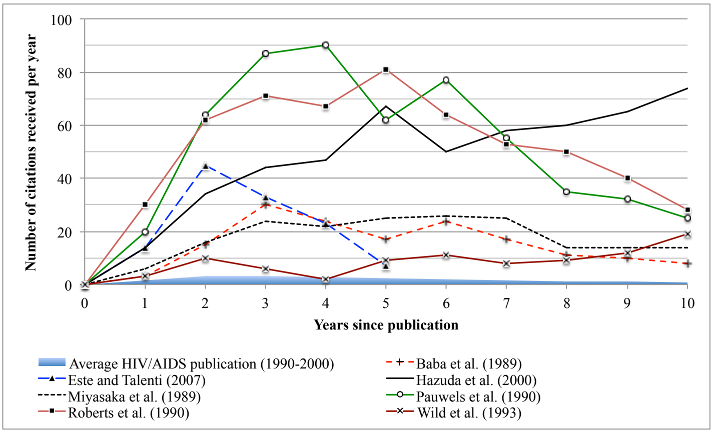

Figure 2.1: Number of citations received per year by the key papers and an average HIV/AIDS publication during the first 10 years after publication

publication as well as the (weighted) average of all HIV/AIDS publications. To compute the weighted average we only included documents that were cited at least once during the period. Each of these seven key papers receives significantly more citations than average. 'Hazuda 2000' is unique in its continued high-volume citation impact, which makes it the only 'citation classic' in its field.

Figure 2.2 exhibits the general trend in the volume of scholarly publications and patent applications related to the class of 'Integrase Inhibitors' medicines. Scholarly publications emerged in 1991. Patent publications took off in 1997.[54] In other words, it took 6 years to utilize and transform scientific knowledge into a patentable technology; even the more efficient 'RDD' approach to develop new drugs is still a complicated and time-consuming trajectory. The graph also shows a steep increase in the number of publications in 2001, after 'Hazuda 2000' went public. A second steep rise occurred in 2005 following the publication of two other highly cited publications by Hazuda: (Hazuda et al., 2004a,b) receiving 12 respectively 14 citations during the first 12 months of their existence. These two belong also to the top 5 % most cited publications in the research field of Integrase Inhibitors.

Table 2.2 shows, for three different citation windows, the relative position of each key paper within the distribution of all HIV/AIDS papers published in the same year. We argue that if a research paper truly represents a scientific breakthrough, is of great relevance for further research, and it is recognized as such by peers working in the same research field, one may expect to see a relatively quick response from peers in the field and steep

[54]Although 'Hazuda 2000' is seen as an important publication it is just a link in the R&D process that started before this publication, and already produced some patents. The publication of this paper might have been intentionally delayed in order not to infringe with ongoing patent applications. The bibliographic data is insufficient to solve this issue.

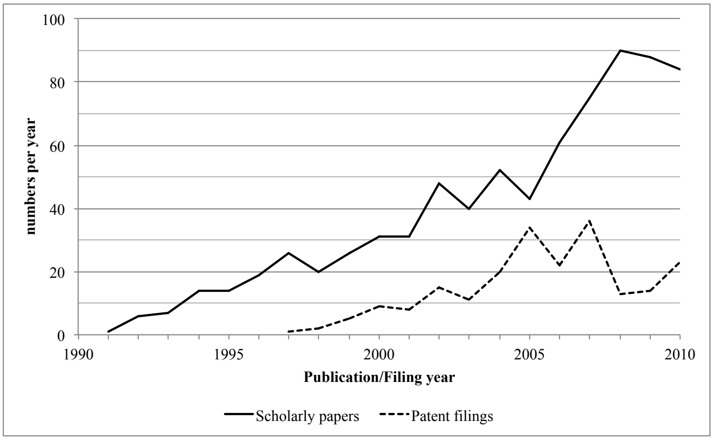

Figure 2.2: Cumulative share of publications as function of the number of citations in the first 12 months after publication

Table 2.2: Relative position of the key papers within the citation distribution of all HIV/AIDS papers published in the same year*

	Citation window		
Publication	1 year	3 years	10 years
Baba et al. (1989)	top 39%	top 7%	top 5%
Miyasaka et al. (1989)	top 25%	top 7%	top 4%
Pauwels et al. (1990)	top 5%	top 1%	top 1%
Roberts et al. (1990)	top 1%	top 1%	top 1%
Wild et al. (1993)	top 33%	top 17%	top 8%
Hazuda et al. (2000)	top 4%	top 2%	top 1%
Este and Talenti (2007)	top 4%	top 2%	not available

* Only citing publications of the document types ‘article’ and ‘letter’ are included

rise in citations to the paper. For early identification of such ‘hot papers’ we restrict our time window to year 1 after publication. So how ‘hot’ was ‘Hazuda 2000’ Figure 2.3 shows the cumulative share of the citing documents as function of the number of citations received during the first 12 months for all HIV/AIDS publications and for all ‘Integrase Inhibitor’ publications in our sample, where the latter is split into citations prior to ‘Hazuda 2000’ (indicated by ‘< 2001’), and citations afterwards until the follow-up publications of Hazuda et al. (2004a,b). While 65 % of the Integrase Inhibitors publications received less than 4 citations during that 1st year, ‘Hazuda 2000’ managed to collect 16 citations, immediately making it one of the most cited in the field. Based on the number of citations received during those first 12 months, four of the seven key papers belong to the top 5 % most highly cited HIV/AIDS publications worldwide.

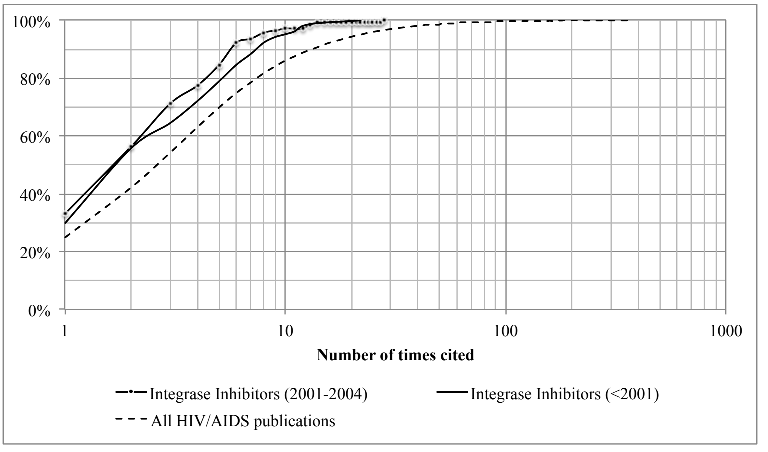

Figure 2.3: Trend of total number of distinct CWTS subject categories citations originate from

2.3.1 Multidisciplinarity of citations

A second assumption relates to the breadth of a publication's citation impact. To help identify a breakthrough we looked at the evolution of the multidisciplinarity[55] We looked at the multidisciplinarity of the citing publications, where multidisciplinarity is defined as the number of different Subject Categories[56]. Figure 2.5 displays the number of additional Subject Categories for each subsequent year. The multidisciplinary citation profile (Figure 2.4) of (Hazuda et al., 2000) is comparable to (Pauwels et al., 1990) and (Roberts et al., 1990). The latter two actually marshal the largest disciplinary spectrum, a result of three waves of expansion in years 3, 6 and 10. What then makes 'Hazuda 2000' so distinctive in terms of the sheer volume of citations? And why its continued high-level citation impact?

2.3.2 Bridging researchers

Of all authors publishing in the field of Integrase Inhibitors only 5.4% are such bridging researchers. Within the patent publications we found that 30.6% of the inventors belong to this select group, which clearly indicates that scientific research and drug development are closely linked in this field. Our analysis of the patents also reveals that all first authors of the seven key papers (Table 2.1) are bridging researchers. De Clercq (2013) commented 'Daria

[55] Porter and Rafols (2009) and Ponomarev et al. (2012) use the related term 'interdisciplinarity' to classify research publications that cover several scientific disciplines.

[56] CWTS uses 251 distinct subject categories based on the 'standard' subject categories of used by Thomson-Reuters in the Web of Science database; 60 of these CWTS subject categories are related to medical sciences and life sciences; the themes range from fundamental biomedical research to clinical medicine. It is expected that citations start to come a increasing and more 'application' oriented range of subject categories as research progresses

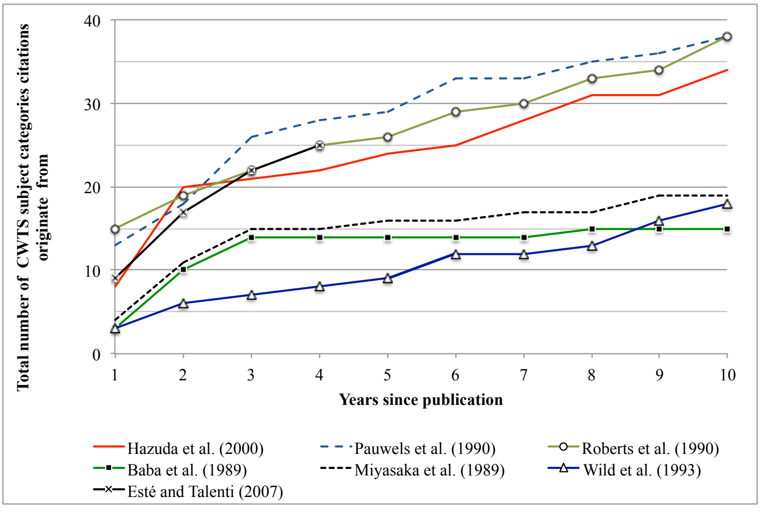

Figure 2.4: Numbers of 'new' subject categories citations originate from added in a year

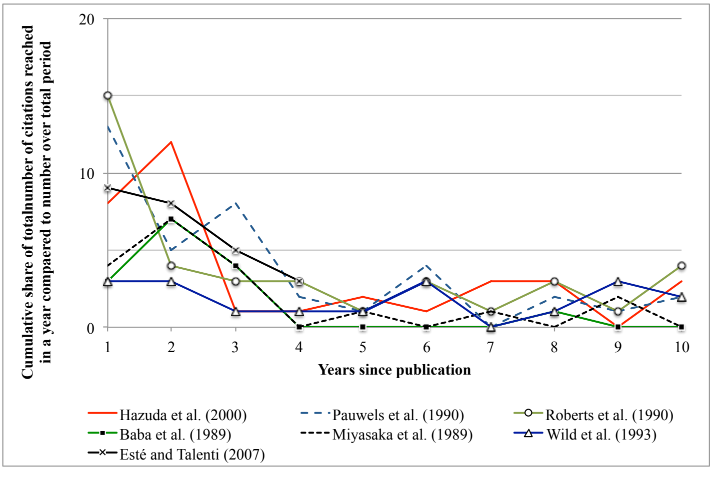

Figure 2.5: Trend of scholarly and patent publications in the field of 'Integrase Inhibitors'

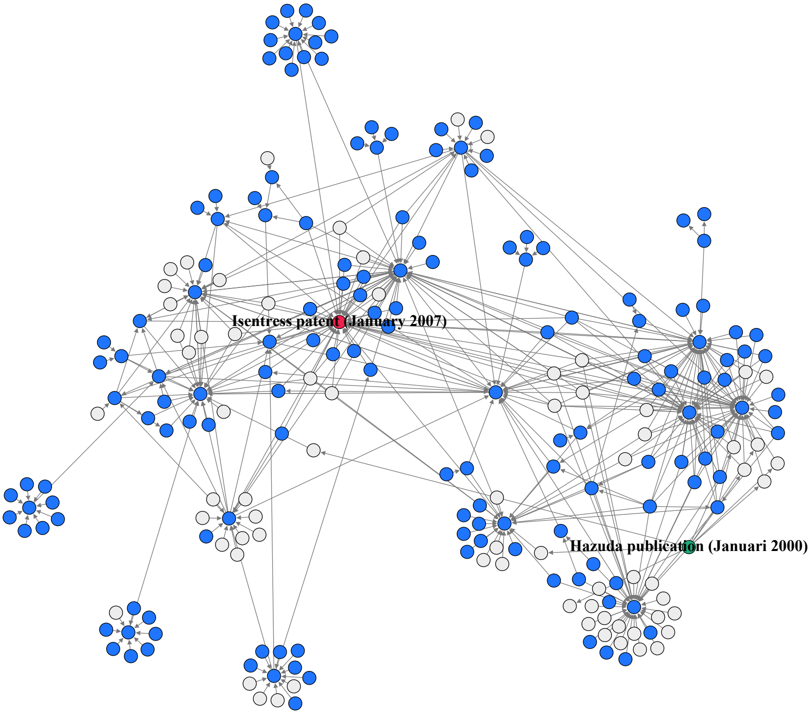

Figure 2.6: Network of patents and scholarly publications connection the Hazuda paper and the Isentress® patent. Blue (opaque discs) discs represent patent publications; white discs represent scholarly publications

Hazuda is a unique example of a bridging researcher, as she is at the origin of the concept (strand transfer integrase inhibitors) as well as the drug [Raltegravir (*Isentress*®)].'

If this paper is indeed such an influential contributor, it should occupy a core node in knowledge creation and utilization networks. Does it? These networks are reflected in citation linkages within and between sets of publications and patents. We compiled a set of patent publications with a citation link to this particular patent. No less than 18 companies and organisations applied for patents that are directly linked to this paper by explicit citation relationships. One of these companies is *Merck & Co* where Hazuda's research was done; one of its subsidiaries owns the *Isentress*® patent. This set of citing patents contains 135 distinct inventors. On the scientific side we identified 1,878 (co-) authors citing 'Hazuda 2000'. All in all, both sources collectively assemble sufficient information to construct the citing-cited network of scholarly publications and patent publications depicted in Figure 2.6. Taking the *Isentress*® patent as a starting point, and following backward citation links to either other patent publications or scholarly publications, we collected cited documents. The search process terminated when a scholarly publication was found or when either the publication date of a publication or the filing date of a patent publication reached the year 1999. Interestingly we found no direct citation links between 'Hazuda 2000' and the *Isentress*® patent. However they do connect in seven cases via intermediate citing-cited links. Six of these intermediates are scholarly publications, including Hazuda et al. (2004a,b), and one is a patent publication. This loose coupling between breakthrough science and patented medicine reflects stepwise processes where 'Hazuda 2000' reports on an initial 'enabling discovery', only one of several building blocks needed to create a new medicine, where Hazuda's follow-up papers in 2004 are more application oriented.

We identified 21 persons as bridging researchers. In Figure 2.7 we present these bridging researchers and the publications related to the development of *Isentress*® they (co) authored. This figure shows the science-technology border in case of the *Isentress*® development based on the citing-cited relations between the publications and links based on (co) authorships relationships. Two clusters are visible; the larger cluster is centred around the 'Hazuda paper'. The three publications by Hazuda (Hazuda et al., 2004a, 2000, 2004b) in this cluster received respectively 16, 12, and 14 citations during the first twelve months after publication.

The other publications in this cluster received three or fewer citations in the same period. This cluster presents the scientific R&D that played a key role in the development of *Isentress*®. Two papers (Hazuda et al., 2004a,b) have seven bridging researchers in common and are linked by a citing-cited relation to the Hazuda paper. Hazuda is (co) author of these three highly cited publications and plays therefore a central role. The papers in the second cluster received only a small number of citations.

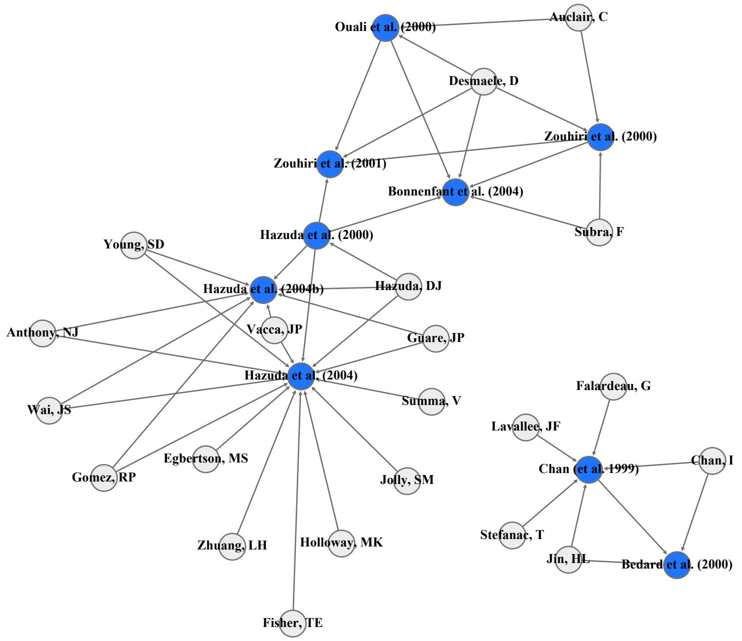

Figure 2.7: Networks of bridging researchers and their publications related to *Isentress*®. Blue (opaque discs) discs represent scholarly publications; white discs represent (co) authors

2.4 Discussion and concluding remarks

This paper develops and suggests a novel analytical framework for early-stage detection of potential scientific breakthroughs. Its applicability hinges critically on several methodological assumptions and technical features related to the empirical data we used and the quality of the databases. In this pilot study we applied several constraints: (1) we restricted our study to only one of the six classes of HIV/AIDS drugs; (2) the focus is on R&D processes cumulating in an approved medicine; (3) a relative short period of time is used for tracking citations; (4) all scholarly and patent publications, linked by citing-cited relations from the initial 'Hazuda 2000' paper to the final *Isentress*® patent, are included; (5) authors and bibliographical links are included insofar they could be identified in the databases.

Using extensive volumes of bibliographic information to identify key research publications and associated 'bridging' researchers, we analysed the R&D dynamics for 'Integrase Inhibitors' and identified what seem to be breakthrough stages in the R&D process. The 'Hazuda 2000' publication is of particular importance according De Clercq (2009b), and the numbers of citations to this paper in the international scientific literature substantiate this claim.

'Hazuda 2000' is receiving, on average, each year 40 citations from scientific research publications and 10 from review articles; of the key papers in the HIV/ AIDS field it is the only 'citation classic', this reflects the fact that the publication serves as reference for new developments, and helps in 'new ways of thinking about a problem' (Hollingworth's definition). The bibliographic data we used indicates that a cascade of R&D activities was made possible as a follow-up of the Hazuda's achievements. Our exploratory study reveals that Hazuda is also one of the researchers that bridge the domains of science and technology. Hazuda is one of several researchers in the case of HIV/AIDS medicines who all made pivotal discoveries as mentioned in the seven key papers in the several research subfields.

In this case study the 'Hazuda 2000' publication is used as starting point to search for alterations in bibliographic data related to breakthroughs in the development of a specific type of HIV/AIDS medicines. According to an expert in this field of research the Hazuda paper describes a breakthrough in the development of medicines in the Integrase Inhibitors group. The main methodological conclusions of our experimental approach are that, under the right conditions, potential scientific breakthroughs in areas of science-based technology might be found by searching bibliographic data sources for highly cited 'key' publications which are (co-) authored by bridging researchers and that show an increasing multidisciplinary impact. These critical conditions depend on: (1) the ability to identify citation links between scholarly publications and patent publications; (2) the maturity of the research area in order to delineate document sets; (3) the coverage of the publications of the research field in the bibliographic databases. The method is therefore — by default — restricted by the information contained in the bibliographic databases we used; we therefore cannot rule out the possibility that we overlooked important features and contributions in R&D processes that are not covered by the databases.

Acknowledging these cautionary remarks, the results of this exploratory case study suggest that indicator-based methodologies are able to, retrospectively, find potential breakthrough research papers relatively soon after their publication in the open scientific literature. The extent to which these general observations and conclusions can be extended to other R&D areas and processes (other than science-driven drug development) remains an open question for now. Further research is needed to apply this approach across wider range of R&D domains, to assess the probability of identifying potential scientific breakthroughs within 1 to 2 years after publication. (1) the ability to identify citation links between scholarly publications and patent publications; (2) the maturity of the research area in order to delineate document sets; (3) the coverage of the publications of the research field in the bibliographical databases. The method is restricted by the information contained in the databases we use and we cannot rule out that our method misses an important discovery because data is not covered by the databases. The extent to which these conclusions can be generalised to other areas of scientific enquiry, R&D processes and technologies other than drug development, remains an open question for now.

2.5 Acknowledgement

The authors kindly thank Prof. (em.) Erik de Clercq for commenting on a concept version of this manuscript.

Table 2.3: Short biography of professor Erik De Clercq

Professor Erik De Clercq

European Inventor of the Year 2008 in the category "Lifetime achievement"

[ ] De Clercq's landmark contributions to biomedical sciences include the development of new antiviral agents for the treatment of viral infections ranging from different forms of herpes to hepatitis B and HIV. In fact, De Clercq and his team are single-handedly responsible for developing some of the original drugs against HIV. They also inspired a switch from monotherapy (treatment with one kind of drug) to the use of up to four agents at a time - the "drug cocktail" that is now the standard treatment approach for HIV. He is best known for opening up the field of so-called "nucleotide analogues", which spawned the first-ever series of broad-spectrum antiviral drugs.

[ ] Since 1972, he has led the Laboratory for Virology at Leuven University's Rega Institute for Medical Research. De Clercq has published over 2100 papers in peer-reviewed scientific journals and has given more than 530 lectures at international congresses, conferences and symposia on various aspects of antiviral chemotherapy. An active member of countless boards and advisory committees, including the World Health Organization (WHO) Expert Advisory Panel on Virus Diseases, De Clercq is addressing the pressing issues of our times, including the threat of an H5N1 "avian flu" pandemic.

Source:

http://www.epo.org/learning-events/european-inventor/finalists/2008/de-clercq.html