

Early stage identification of breakthroughs at the interface of science and technology: lessons drawn from a landmark publication

J. J. Winnink · Robert J. W. Tijssen

Received: 20 November 2013 / Published online: 4 October 2014
© Akadémiai Kiadó, Budapest, Hungary 2014

Abstract Certain scholarly publications or patent publications may signal breakthroughs in basic scientific research or radical new technological developments. Are there bibliographical indicators that enable an analysis of R&D dynamics to help identify these ‘local revolutions’ in science and technology? The focus of this paper is on early stage identification of potential breakthroughs in science that may evolve into new technology. We analyse bibliographic information for a typical example of such a breakthrough to pinpoint information that has the potential to be used as bibliographic indicator. The typical example used is the landmark research paper by Novoselov et al. (Science 306(5696): 666–669, 2004) concerning graphene. After an initial accumulation of theoretical knowledge about graphene over a period of 50 years this publication of the discovery of a method to produce graphene had an immediate and significant impact on the R&D community; it provides a link between theory, experimental verification, and new technological applications. The publication of this landmark discovery marks a sharp rise in the number of scholarly publications, and not much later an increase in the number of filings for related patent applications. Noticeable within 2 years after publication is an above average influx

Electronic supplementary material The online version of this article (doi:[10.1007/s11192-014-1451-z](https://doi.org/10.1007/s11192-014-1451-z)) contains supplementary material, which is available to authorized users.

J. J. Winnink (✉) · R. J. W. Tijssen
Centre for Science and Technology Studies (CWTS), Leiden University, Wassenaarseweg 62A,
2333 AL Leiden, The Netherlands
e-mail: winninkjj@cwts.leidenuniv.nl

J. J. Winnink · R. J. W. Tijssen
Leiden University Dual PhD Centre The Hague, P.O. Box 13228, 2501 EE The Hague,
The Netherlands

J. J. Winnink
Netherlands Patent Office, P.O. Box 10366, 2501 HJ The Hague, The Netherlands

R. J. W. Tijssen
DST-NRF Centre of Excellence in Scientometrics and STI Policy, Stellenbosch University,
Stellenbosch, South Africa

of researchers and of organisations. Changes in the structure of co-citation term maps point to renewed interest from theoretical physicists. The analysis uncovered criteria that can help in identifying at early stage potential breakthroughs that link science and technology.

Keywords Breakthrough · Early stage · Graphene · Science–technology interface · Weak signals

Introduction

Martin (1995) states ‘[] Confronted with increasing global economic competition, policy-makers and scientists are grappling with the problem of how to select the most promising research areas and emerging technologies on which to target resources and, hence, derive the greatest benefits []’ Identifying breakthroughs in science at early stage could aid in solving this selection issue, and help in timely prioritizing R&D resources by for instance policy-makers, research organisations, or companies. The focus of our research is on finding indicators in bibliographic data that would enable us to conclude at early stage if a discovery is to be considered a potential breakthrough discovery at the interface¹ of science and technology. We study well-known cases generally recognized by experts as breakthrough discoveries in science. In this study the focus is on the publication Novoselov et al. (2004) for which the Nobel Prize for Physics 2010 was awarded to Konstantin S. Novoselov and Andre K. Geim for ‘succeeding in producing, isolating, identifying and characterizing graphene’ (Nobel Prize Physics 2010). We will refer to the paper in which this discovery was published as the ‘Novoselov paper’. Graphene² seems to be an appropriate case in point to study because this paper acted as a bridge between basic science and applied science, and therefore marks the beginning of technological applications. Prior to this seminal paper graphene was not available as an independent material and therefore it could not be studied in experiments or used as the basis for new technological applications.

The properties of sheets of multi-atom thick layers of graphite had already been studied for a considerable period, for instance Wallace (1947) and Boehm et al. (1962). These theoretical studies focused on the physical aspects of graphite layers (sheets). Over time, the theoretical studies uncovered several properties resulting in the expectation that ‘freestanding graphene’ would be a material with a high potential for novel and unforeseen technological applications. Given the state of graphene research at that time, the ‘Novoselov paper’ is to be considered a landmark event that transformed theoretical scientific knowledge into new technological applications. The ‘Novoselov paper’ linked theoretical considerations about graphene and its properties to experimental verification. The

¹ Interface = ‘a point where two systems, subjects, organizations, etc. meet and interact’ (Oxford English Dictionary).

² Graphene is a material that in its purest form is a one-atom thick layer of carbon foil. In graphene, carbon atoms are densely packed in a regular (hexagonal) pattern. These hexagonal patterns are the basic structural element of related materials, including graphite, charcoal, carbon nanotubes and fullerenes. Graphene can also be thought of being an indefinitely large aromatic molecule. The hexagonal arrangement of carbon atoms in a single layer results in quantum mechanical effects that manifest at macro scale as extraordinary properties. Its interactions with other materials and with light in combination with its inherently two-dimensional nature produce unique properties, such as the bipolar transistor effect, ballistic transport of charges and large quantum oscillations. High-quality graphene is strong, light, almost transparent and an excellent conductor of heat and electricity.

discovery presented in this paper proved that graphene could exist in ‘free form’,³ and provided a method to produce sheets of graphene, albeit of limited size. The resulting upswing in graphene R&D not only led to inventions and interesting applications, it also stimulated the application of modern theoretical physical concepts, much more advanced than those available to Wallace in 1947. The application of these new theories resulted in the discovery of previously unanticipated properties (Frenken 2013). This renewed attention from theoretical physicists is reflected in the evolution of the cognitive structure as presented in graphene related publications. In view of the extended and expanding theoretical knowledge base, and the expected economic potential for graphene-based technologies we expect patent applications to appear soon, years rather than decades, after the breakthrough.

We searched for possible changes visible in bibliographic data related to the ‘Novoselov paper’ that can be used as indicators to help in the early stage identification of a publication as a potential breakthrough, and that can reveal the characteristics of this potential breakthrough. In this analysis we focus on the role the ‘Novoselov paper’ had on graphene-related developments, and the diffusion of the knowledge presented among researchers. We use bibliographic data to answer the question ‘Did this landmark paper by Novoselov et al. give rise to a new theoretical framework to explain the properties of graphene?’ We conclude that the ‘Novoselov paper’ describes a ‘charge’⁴ breakthrough, and propose indicators that can be extracted from bibliographic data, and can be used to assist in identifying potential breakthroughs at early stage.

Justification for this study

Research on emerging scientific fields and technology forecasting is conducted from various perspectives, and with various goals in mind. In a preliminary study Adams (2005) concludes that for six research categories across the life and physical sciences there is a significant correlation between early (years 1–2) citation counts and later (years 3–10) citation counts. Andersen and Borup (2009) focus on improving the impact of foresight strategy processed in national research councils and research programs. A foresight model based on an alternative version of the Delphi-AHP approach to detect key areas in the Information Technology is proposed in Bañuls and Salmeron (2008). Bettencourt et al. (2008) shows that the evolution of emerging fields in scientific disciplines is well described by population contagion models. In Breiner et al. (1994) the outcomes of a joint Japanese German Delphi study are analysed to explain possible differences in the outcomes, and to understand the cultural influences on technology assessment. Chen et al. (2009) model transformative discoveries focussing on connections across structural holes in network representations of scientific knowledge in scientific discovery. Coates et al. (1994) reviewed and commented for ‘Project 2025’ all the science and technology forecasts for 1970 and any time forward they could find. Technology Forecasting (TF) and the changes TF underwent over the years are discussed in Coates et al. (2001). Integrating multiples techniques to forecast emerging technologies by using bibliometric data and patent analysis as sources for historical data is discussed in Daim et al. (2006). Hand (2009) addresses the challenges and difficulties involved in the use of the vast data sets used in forecasting.

³ ‘Free form’ is to be understood as a one-atom thick layer of graphene placed on a substrate. The only function of the substrate is to support and fix the graphene layer. The graphene in the layer behaves as if it were freestanding.

⁴ See the short description of Koshland’s Cha–Cha–Cha theory in footnote 6 on p 4.

Julius et al. (1977) present an organisational model in which expert knowledge is used to judge at an early stage whether a discovery is expected to be of importance for new technology that will affect R&D in the pharmaceutical sector. Leydesdorff and Rafols (2011) use geographical and cognitive diffusion to analyse emerging topics. A model for a foresight process to identify scientific developments of strategic importance is presented in Martin (1995). Meade and Islam (1998) provides an overview of 29 different mathematical models that can be used in technology forecasting, and conclude ‘[] The straightforward policy of identifying an appropriate model and then using it to generate forecasts was shown to be difficult, if not impossible, to put into practice. []’. Mishra et al. (2002) proposes the mapping on a common scale of characteristics of technology and techniques to match technique with the technology to be forecasted. Ponomarev et al. (2012, 2014) uses citation patterns in combination with statistical modelling to search and trace developments in potential breakthrough publications. A combination of a fuzzy Delphi method, analytic hierarchy process (AHP), and patent co-citation approach (ACP) to select emerging technologies is presented in Shen et al. (2010). Small (1977) concludes that the citation picture of a specialty’s development in science matches the perception of the specialists in the field. Tu and Seng (2012) propose two indexes, the novelty index (NI) and the published volume index (PVI) in order to detect emerging topics. Yoon and Park (2007) suggests combining morphology analysis, conjoint analyses and citation analysis of patent information to forecast new technologies. In Young (1993) nine different growth curve models are fitted onto various data sets in an attempt to determine the growth curve model that achieves the best forecasts.

We are looking for indicators in bibliographic data that signal, at early stage, the moment a publication of a potential breakthrough in science occurs. The precise point in time a breakthrough occurs often cannot, even in retrospect, be precisely isolated on a time line. We argue, however, that it is possible to gauge major breakthroughs retrospectively, within limits, by using bibliometric information to identify related landmark publications. Koshland (2007) differentiates in his *Cha–Cha–Cha–theory*⁵ scientific discoveries in three distinct classes on the basis of the nature of the discovery in relation to already existing scientific knowledge. Hollingsworth (2008) defines ‘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’. Our focus is on major role-changing discoveries, which we will call *breakthroughs*. Combining the ideas of Koshland and Hollingsworth enables us to differentiate between breakthrough discoveries and small advance discoveries, and opens up the possibility to classify breakthrough discoveries into the three categories defined by Koshland. In this study we interpret the aforementioned definition from Hollingsworth ‘[...], which opens up new ways for further physical research’. Kuhn (1962, Ch. VII, VIII) argues that crisis and theory change go hand in hand, and concludes that a crisis does not immediately lead to the replacement of the failing theory by a new one. ‘Challenge’ discoveries can be seen as revolutionary science in Kuhnian sense, and proliferate slow as result of reluctance to supersede an existing theoretical framework with a new one. ‘Charge’ discoveries on the other hand can almost instantaneously lead to new R&D

⁵ The three different types of scientific discovery according to Koshland (2007) are:

Charge discoveries, not necessarily simple, solve scientific problems that are quite obvious and fit within the at that moment existing theoretical framework.

Challenge discoveries are a response to an accumulation of facts or concepts that are unexplained by or incongruous with scientific theories of the time, ‘paradigm shifts’ (Kuhn 1962).

Chance discoveries are those that are often called serendipitous. Van Andel (1994) denotes discoveries of this type ‘unsought findings’.

activities as they solve scientific puzzles for which no immediate changes in the existing theoretical framework are needed. We therefore expect that a discovery leaves behind a pattern in bibliographic data that is specific for the type of this discovery.

The identification of breakthrough moments is relevant because they signal possible breaches (focus shifts) or even turning points in the R&D system. Martin (1995) points at the problems policy-makers and scientists have in selecting the most promising research areas and emerging technologies. Identifying at early stage publications of potential breakthrough discoveries could assist this selection process. Policy makers in industry of government could adopt their strategy based on the notion a possible breakthrough exists. We consider the moment patent applications that are closely connected to a research front start to appear signal that basic scientific knowledge from this research front becomes to be used for ‘real world’ applications. Geim and Novoselov (2007) wrote in the abstract of their publication ‘The rise of graphene’ “[...] *the graphene ‘gold rush’ has begun.*” The properties of graphene, forecasted by theoretical physical research, could only be exploited if freestanding graphene was isolated. The ‘Novoselov paper’ marks the start of the transformation of basic science into technological applications. This publication also leads to the application of more modern theoretical concepts than those available to Wallace and Boehm, which on their turn resulted in the prediction of other not yet foreseen properties.

We consider science-based innovations as innovations that are only made possible by newly acquired scientific knowledge. For science-based innovation systems a long period between the development of the basic theoretical concepts and the first technological applications seems to be the rule rather than the exception. A number of researchers, including Jewkes et al. (1958), Isenson (1969), and Grupp and Schmoch (1992), conclude that the diffusion of basic scientific knowledge into technological applications often takes dozens of years. The accumulated theoretical knowledge is often released precipitately as a result of a single pivotal event (scholarly publication). Graphene is an example of such a science-based innovation. We see the ‘Novoselov paper’ as a pivotal publication and consider the graphene case to be particularly suited to our search for bibliographical indicators. The fact that there is one single event, the ‘Novoselov paper’, that describes the breakthrough discovery, results in a well-defined point in time that marks this event. According to the Nobel Prize committee (Nobel Prize Physics 2010) this paper clearly marks the ‘graphene’ breakthrough. We discussed the ‘Novoselov paper’ and its effects on graphene research with professor Frenken, one of the researchers active in graphene research (see Table 5 for background information). We argue by analysing bibliographic data, applying Koshland’s criteria, and using Hollingsworth definition that the ‘Novoselov paper’ can be considered a typical example of a charge breakthrough. This conclusion is in line with information from the Nobel Prize committee, and the results of the discussion with Frenken.

Hypothesis and research question

The accumulation of basic scientific knowledge of graphene took place over a period of almost 60 years. The ‘Novoselov paper’ showed a way to produce small sheets of free-standing graphene, in this way graphene became available for experimentation. We therefore expect the number of scholarly publications on graphene as well as the number of graphene related patent applications to rise significantly and simultaneously after a short ‘incubation’ period. A significant rise is anticipated because the breakthrough event removes the barrier for usage of the accumulated theoretical knowledge for applied scientific research. The expected upswing should, at first, coincide with the increase in the

number of citations of the ‘Novoselov paper’ well before the Nobel Prize was awarded. We furthermore expect to find evidence in the bibliographic data authors and organizations not previously engaged enter the graphene arena.

Our hypothesis ‘Bibliographic data contains information that enables to identify and characterize at early stage publications of potential breakthroughs at the interface of science and technology’ is to be answered by the research question ‘Can we identify indicators in bibliographic information that can help in identifying and characterizing at early stage publications that describe a potential breakthrough at the interface of science and technology?’

To answer the research question the focus is on the following criteria to conclude if a typical publication contains a potential breakthrough that might result in a new technology. The criteria consist of two types, criteria 1 and 2 focus on very properties of a discovery, and the criteria 3, 4, and 5 on the impact of the discovery on the scientific research system:

1. The scientific discovery has the potency to evolve into a new technology
2. If the discovery introduces a new theoretical framework it is classified as a ‘challenge’ discovery, otherwise it is classified as a ‘charge’ discovery. Characterisation of a breakthrough is important as we think that breakthroughs leave behind specific patterns in bibliographic data that reflect the characteristics of the type of the breakthrough.
3. The publication is highly cited from the moment it is published. In this study we exclude publications with a delayed interest, the ‘sleeping beauties’ (Van Raan 2004)
4. The discovery leads to renewed interest for the theory underlying the discovery. This renewed interest is driven by the outcomes of experimental research related to the discovery
5. The discovery results in an above average influx of researchers and organisations new to the research field; the potential breakthrough gains much attention of scientist working in the same or adjacent areas, who want to participate in this new direction. The manner in which knowledge presented in a publication diffuses among scientist and organisations is an indication of the type of a discovery.

Hollingsworth (2008) provides a definition of a breakthrough, and in combination with Koshland’s Cha–Cha–Cha-theory this should make it possible to distinguish between charge discoveries and challenge discoveries. Consulting an expert on the subject helps in validating the findings, and especially to conclude if a paradigm shift emerged from the discovery.

We analyse bibliographic information related to the ‘Novoselov paper’ to learn if these criteria can be used to base indicators on that can be of help in early stage identification and characterization of potential breakthroughs at the interface of science and technology.

Data and information sources

Scholarly publications

We used Thomson-Reuters Web of Science database (WoS) as source for bibliographic information on scholarly publications. The WoS contains, according to Thomson-Reuters website, over 55 million records with bibliographic data for scholarly publications from journals (>18,000), conference proceedings, and books in the sciences, social sciences, and arts and humanities. For a detailed analysis of the evolution of graphene research we collected, from the in-house version of the WoS (TR/CWTS WoS), for the relevant

publications terms occurring in titles and abstracts to create co-occurrence maps. We used the date of publication to put scholarly publications on the time scale. The publication date used in the WoS is the ‘date’ on the front cover of the journal.⁶

As long as graphene was not available as freestanding substance research was closely linked to research on graphite; graphene was hardly seen as a separate field of research. To select relevant publication from the WoS we used the topic⁷ ‘graphene’. This approach resulted in a data set containing information for 18,499 scholarly works⁸ for the period 1990–2012.

Patent publications

We used the April 2012 version of PATSTAT⁹ to collect the relevant patent data. To select patent publications we relied on graphene-specific patent classification codes assigned to the patent publications. The use of patent classification codes is to be preferred over keyword searches; as skilled patent examiners assign these classification codes to patent documents. Widely in use is the International Patent Classification system (IPC). Appropriate classification codes from the IPC are assigned to all patent applications. The IPC however does not contain graphene specific codes to date. A more fine-grained classification system compatible with the IPC was in use at the European Patent Office (EPO) until 2013.¹⁰ This system called the European Classification system (ECLA) was accompanied by an even more specific classification system called In Computer Only (ICO). Both ECLA and ICO contain several graphene-specific classification codes.¹¹ Patent examiners at the EPO assigned the moment a patent application was processed appropriate ECLA and ICO codes.

Graphene is considered a material with a bright technological future, and is expected to generate large revenues. Therefore we expect that the vast majority of graphene related inventions would end up as patent application at the EPO as one of the major patent offices in the world. The drawback in using ECLA and ICO codes is that no such code is assigned to a patent application until the moment it is processed by an examiner at the EPO. We compensate for this systemic error by additional including in our data set bibliographic information for patent documents of which the abstract or the title contains the phrase ‘graphen*’. Other search terms were tested but inspection of the results learned that these did not reveal additional relevant publications without introducing non-relevant documents. We are confident that the dataset covers the patent publications in an adequate way.

⁶ A more accurate approximation for the date of a discovery would be the date a manuscript was submitted for the first time to a publisher. This submission date is not available in the databases and could therefore not be used. The publication itself is not necessarily the first attempt to publish the discovery resulting in an even more imprecise time stamp for the discovery.

⁷ According to Thomson-Reuters information terms for a topic are extracted from the title, the abstract, the author keywords of a publication, and from KeyWords Plus[®]. ‘KeyWords Plus are words and phrases harvested from the titles of the cited articles.’

⁸ In this study we focus on publications containing original scientific work, and therefore only publications of the WoS types ‘article’ and ‘letter’ are used in the analysis.

⁹ EPO Worldwide Patent Statistical Database (PATSTAT).

¹⁰ The Cooperative Patent Classification (CPC), which is the result of cooperation between the EPO and US Patent and Trademark Office (USPTO) replaced in 2013 the ECLA, ICO, and the classification of the USPTO (US Patent Classification).

¹¹ We used ECLA and ICO codes C01B31/04H, H01F41/42, H01F42/44, H01F10/00C, H01L29/16G, M01B204/00, M01B204/02, M01B204/06, M01B204/20, T01F10/00C, T01L29/16G.

An intrinsic effect of the patent system is that inventions often lead to multiple equivalent patent publications in several jurisdictions.¹² To count ‘inventions’ we grouped equivalent patent documents together in DOCDB-patent families; such a patent family represents one invention. The earliest filing date of all patent applications within a patent family is used as approximation for the date of the invention. The resulting data set contained 2,083 patent families applied for in the period 1990–2011. The data set does not fully cover patent filings from Autumn 2010 onwards due to the secrecy period of 18 months, that is part of the patenting procedures, and the use of the April 2012 release of PATSTAT.

Citations in patents reference other patents as well as other, for instance scholarly, publications. Especially patent publications published by the United States Patent and Trademark Office (USPTO) contain a substantial list of these front-page¹³ citations. Other patent offices usually publish only citations that occur in ‘search reports’; the citations mentioned in these reports constitute only a small portion of the citations in the patent documents. Especially in the case of technologies considered very promising, like graphene, there is often at least one publication from the USPTO present in the patent family. Of all DOCDB families in the PATSTAT database approximately 20 % contain at least one USPTO publication; for the graphene patent publications in our dataset this is the case in 41 % of the patent families. In the analysis all relevant bibliographic data were taken into account. The use of patent families solves, at least in part, the problems with references in patent publications to scholarly literature.

The graphene breakthrough

Section “[The evolution of the graphene field](#)” focuses on the evolution of the graphene field, and uses absolute numbers of scholarly publications and patent applications to visualize it. Breakthrough publications are expected to stimulate research, and therefore result in a steep increasing number of related publications. To analyse the impact of the ‘Novoselov paper’ over time the share of scholarly publications and patent publications citing this publication are calculated for each year. It is expected that breakthrough publications be among the publications with a high citation count. In section “[Has the ‘Novoselov paper’ been a highly cited paper from the moment it was published?](#)” we compare the number of citations the ‘Novoselov paper’ received during the first 12, 24 and 36 months with the numbers received by graphene publications published in 2004 and 2005, and by the publications from Volume 306¹⁴ of *Science*. According to Koshland ‘Challenge’ discoveries lead to changes in the theoretical framework. The publications that cite de ‘Novoselov paper’ were classified into six categories; ranging from discovery science to applied science. We calculated (“[Did the ‘Novoselov paper’ introduce a new paradigm?](#)” section) for each year in the period 2005–2012 for every category the absolute numbers of publications, and the share the total number of publications in that year. A drop in the number of publications in the category ‘discovery science’ while at the same time the numbers for the other categories increase is an indication that the scientific knowledge from the publication becomes used in applied areas. The balance between basic research and applied research is discussed in section “[Balance between basic research and applied](#)

¹² Patent rights are nowadays national rights granted by a sovereign state.

¹³ In the US the list of citations appear on the front-page of the patent publication.

¹⁴ The volume in which the ‘Novoselov paper’ was published.

research”. The publication of a breakthrough discovery stimulates researchers to become active in the new area, and start citing in their publications the breakthrough publication. In section “[The inflow of researchers and organisations in graphene research as result of the ‘Novoselov paper’](#)” the first occurrence of first authors, and their affiliations, that cite the ‘Novoselov paper’ in their publications is used to approximate this influx. The possible effect of the ambiguity of names is discussed in section “[Influence of ambiguity of author and organisation names](#)”.

The evolution of the graphene field

Figure 1 presents trends in absolute numbers for graphene-related scholarly publications, publications citing the ‘Novoselov paper’, and for patent families. After the publication of the ‘Novoselov paper’, the ‘graphene field’ shows an upswing in R&D visible as a sharp increase in the number of publications later accompanied by a remarkable rise in patent applications. The figure shows furthermore the number of scholarly publications citing the ‘Novoselov paper’. The share of these publications topped in 2009 at 49 %, and since declines as is shown in Table 1.

It is known that certain events, such as the award of a Nobel Prize, can have a significant influence on the number of citations that publications of laureate authors receive. In this study this effect plays no role as the focus is on citations made well before the Nobel Prize Physics was awarded to Geim and Novoselov in 2010. The last two columns in Table 1 show that citations in patent publications to the ‘Novoselov paper’ are present since 2006.¹⁵ The patent data for 2010 and 2011 is incomplete as result of the secrecy period, which is part of the patenting process. Citations made by patent applicants are comparable in nature with citations in scholarly publications. In most cases the applicant of a patent, as is shown in this table, added the citations to the ‘Novoselov paper’. These citations indicate the existence of a link between science (scholarly publications) and technology (patent publications). The number of patent families citing the ‘Novoselov paper’ is low in comparison with the number of ‘graphene’ patent families that have been filed. The discovery in the ‘Novoselov paper’ showing the method to isolate graphene as a substance so it could be used in experimental research influenced, directly or indirectly, much graphene related research. The low number of citations in patent applications suggests that to the majority of graphene related inventions the discovery is not seen to be of particular significance, and as patent applicants are not obliged to cite previous literature in their patents it is therefor not cited.

Further evidence for the limited importance of this paper for patenting, is given by the fact that the patent families in the dataset contain in total 2,462 references to scholarly publications; only a small share (3 %) are citations to the ‘Novoselov paper’; these citations originate predominantly from patent applicants. Patent examiners usually only cite publications that are relevant to the granting decision they have to take, and document them in search reports. By using patent families of which, at least in this case, a large number (41 %) contain an USPTO patent publication a bias towards citations originating from the patent examiners is largely prevented. The reason for this is that in the US patent applicants are obliged to cite all relevant patent and non-patent literature on which their invention is based. The large influence of the ‘Novoselov paper’ on the development of graphene related R&D is shown by the share of scholarly publications citing it. The share

¹⁵ Inspection of the patent file for the patent application from 2004 citing the ‘Novoselov paper’ leads to the conclusion that at a later stage in the patenting process this citation was added; most probably in 2006.

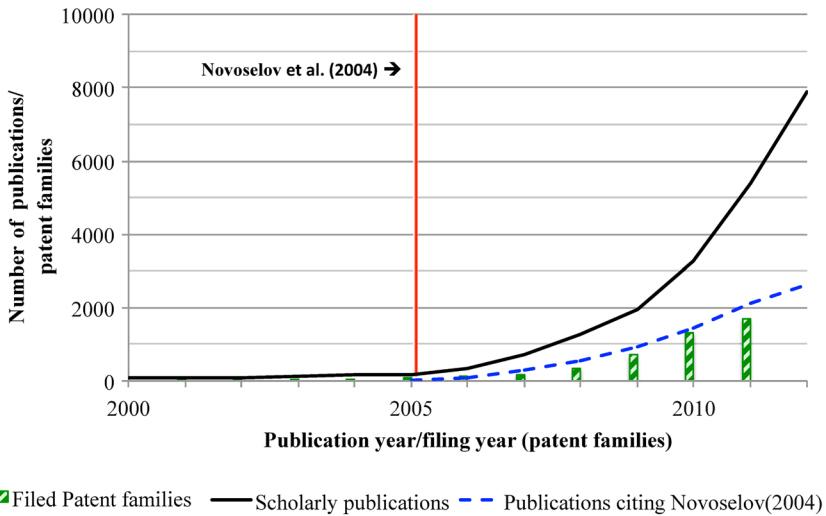


Fig. 1 Scholarly publications and DOCDB patent families related to graphene (2000–2012)

of graphene publications citing that paper topped in 2009 with 46 %, and decreased since; the absolute number of citations is still rising. As the graphene research field matures research on graphene digresses from the landmark publication.

We conclude that the ‘Novoselov paper’ is at the interface of science and technology as it is referenced in patent publications. The direct influence on the development of graphene technology seems to be minor as the paper is cited in only a small number of graphene related patents. This should not come as a surprise as the discovery in the ‘Novoselov paper’ is in producing, isolating, identifying and characterizing graphene. Inspection of the patent publications shows that most inventions focus on the application of graphene and not on producing graphene.

Has the ‘Novoselov paper’ been a highly cited paper from the moment it was published?

In order to assess whether the ‘Novoselov paper’ is a highly cited paper we compared the number of citations it received with the numbers for all other graphene publications¹⁶ from 2004 and 2005. We computed the share of the publications that received a certain number of citations within 12, 24, and 36 months after publication.

The results are shown in Fig. 2. The Y-axis in this figure represents the share of the publications that received at most the number of citations of the corresponding point on the X-axis. The position of the ‘Novoselov paper’ is indicated on each of the curves by opaque (red) disks. This figure and also Table 2 illustrate that among the graphene publications from 2004 and 2005 the ‘Novoselov paper’, from the moment it was published, is one of the most cited. Based on the number of citations received in the first 36 months (Table 2) the ‘Novoselov paper’ belongs to the top 1 % of publications from Volume 306 of Science, the volume it appeared in. The publication in Science, an esteemed journal, with its wide

¹⁶ We included only publications with document types research articles and letter, from 2004 to 2005.

Table 1 Scholarly publications (articles, letters), patent filings and citation to the ‘Novoselov paper’

Year	Scholarly publications on the topic graphene			Graphene DOCDB patent families		
				Citations to the ‘Novoselov paper’		
		Citing the ‘Novoselov paper’	Share of scholarly publications citing the ‘Novoselov paper’		Made by the patent applicants	Made by the patent examiners
2000	81	.	.	29	.	.
2001	90	.	.	39	.	.
2002	104	.	.	38	.	.
2003	130	.	.	52	.	.
2004	151	.	.	49	1 ^a	0
2005	182	10	5 %	46	0	0
2006	323	81	25 %	88	2	0
2007	703	272	39 %	142	1	1
2008	1,270	553	44 %	318	9	2
2009	1,933	880	46 %	650	8	0
2010	3,246	1,402	43 %	614 ^b	13	1
2011	5,363	2,024	38 %	226 ^b	2	2
2012	7,885	2,543	32 %	.	.	.

‘.’ That data is unavailable

^a This citation was most probably added in 2006, see also footnote 16

^b The data is incomplete as result of the secrecy period that is part of the patenting procedures. Patent families filed in 2010 are for a greater part covered by the PATSTAT release used; the patent families filed in 2011 are covered for only a small part

audience might have contributed to the fact that it received from the moment it was published a considerable number of citations. By the end of the year 2013 the ‘Novoselov paper’ was cited 11,623 times in research articles, and 575 times in review¹⁷ publications. We furthermore compared the number of citations this landmark paper received with the number of citations received by the 318 publications published in the same volume (306) of Science. The last column of Table 2 shows the results. Based on citation counts for the first 24 months the Novoselov paper was in all cases among the top 5 % of publications.

Did the ‘Novoselov paper’ introduce a new paradigm?

The ‘Novoselov paper’ did not introduce a new paradigm as was implicitly mentioned in (Nobel Prize Physics 2010), and explicitly in (Frenken 2013). To check if bibliographic data contains information to answer this question we used the classification scheme presented Table 3 to assign one of six science levels to publications.¹⁸ The methodology and background of this classification are presented in Tijssen (2010). The classification ranges from ‘Discovery Science’ to ‘Applied science’. Individual publications inherit the science

¹⁷ In this paper the classification of a publication is the classification as it appears in the Web of Science database.

¹⁸ The science levels are assigned to journals on the basis of the addresses of the authors of the publications in a journal. Based on the addresses the affiliation of an author is classified as university, hospital, or company. The distribution of the three classes determines the science level assigned to a journal.

level assigned to the journal in which they are published. Only publications with science levels ‘Discovery science’, ‘Industrially relevant science’, ‘Science-based technological development’, and ‘Industrial/medical development’ appear in the document set. The class ‘Discovery science’ is assigned to the ‘Novoselov paper’ as this is the science level assigned to the Science journal.

To answer the question “Did the ‘Novoselov paper’ introduce a new paradigm?” we analysed the evolution of the number of publications citing the ‘Novoselov paper’ that are assigned to the different science levels. ‘Discovery Science’ is the category that encompasses theoretical research. Figure 3 shows the absolute numbers of publications citing¹⁹ the ‘Novoselov paper’ disaggregated into science levels. Publications in the category ‘Industrially Relevant Science’ dominate the picture from the beginning.

In Fig. 4 the shares of the different science levels of the citing publications per year are shown. The share of publications citing the ‘Novoselov paper’ from the category ‘Science-based technological development’ increased from 2005 until reaching a maximum in 2008; from then on it gradually declines. Since 2005 the share of publications in the ‘Discovery science’ category citing the ‘Novoselov paper’ increases at a slow pace; in 2007 it dropped from 20 to 17 % and is increasing since. This is in accordance with the notion that theoretical physicists became again interested in graphene and started to apply more recent theoretical approaches. It is also in line with the discussion in the next section (“[Balance between basic research and applied research](#)”). From the beginning the majority of the publications belong to the category applied science, especially ‘Industrial Relevant Science’. In case the ‘Novoselov paper’ would have introduced a new theoretical framework we expect, at least during period immediate following the publication, dominance in absolute and relative numbers of publications from the ‘Discovery Science’ category citing it; this is not the case. We therefore conclude that the bibliographic data shows evidence that the ‘Novoselov paper’ did not introduce a new theoretical framework.

Balance between basic research and applied research

To analyse the evolution of graphene related research we extracted terms from the abstracts and titles of all graphene related papers in the document set. Winnink and Tijssen (2014) visualizes for the period 2005–2012 the cognitive structure of graphene research in every year by creating co-occurrence maps. For each year the co-occurrence maps are the result of combining terms from the abstracts and titles of all graphene publications up-to that year. Development and shift of focus in graphene research should be visible as changes in these co-occurrence maps²⁰ over time. We used the VOSviewer²¹ for the visualisation. For all pictures the same settings for the visualisation parameters were used to assure that if changes could be observed these changes are caused by alterations in the cognitive structure, and are not the result of different parameter settings. A dichotomy in two areas of more or less equal intensity becomes manifest in 2007. From 2007 onwards the pictures show a more intense area on the right hand side that is related to basic theoretical research,

¹⁹ In order to get reliable results (Tijssen 2010) not all journals, and hence not all publications, have a science level assigned. This results in numbers of publications that are somewhat lower than those presented in Table 1.

²⁰ These pictures were presented before at the STI 2012 conference (Winnink 2012).

²¹ <http://www.vosviewer.com>.

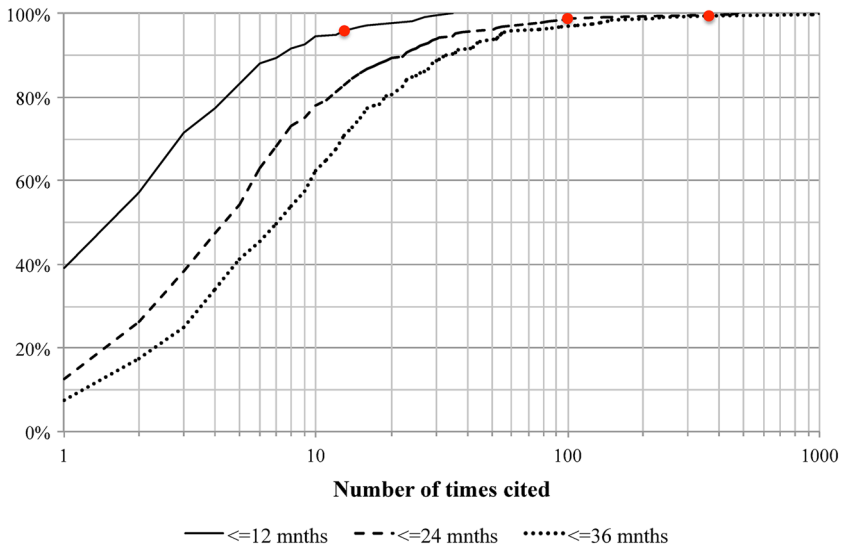


Fig. 2 Cumulative distribution of graphene related publications from 2004 and 2005 based on the number of citations received during 12, 24, and 36 months after publication. The opaque (red) disks indicate the ‘Novoselov paper’

and a second area composed of terms related to applied scientific research on the left. We expect this constellation in which the two areas are in ‘balance’ to be meta-stable; as time proceeds and graphene becomes more mature we expect this balance to change in the direction of applied scientific research. The pictures shows that applied scientific research became more important but at the same time basic (theoretical) research did not diminish. The data suggests that discovery science supports the further development of graphene. This observation is in line with the information presented in Fig. 4, and the outcomes of the discussion with Frenken (“Expert opinion” section) on the question ‘What causes the dichotomy between applied science and theoretical science that becomes apparent in 2007–2008 in the co-occurrence maps of terms from the abstracts and the titles of graphene publications, and that evolves into a (meta) stable configuration?’

The evolution shown in the pictures is also in line with Frenken (2013) stating that the ‘Novoselov paper’ was at the basis of research that proved existing predictions of the properties of graphene to be true, and stimulated the application of more recent theoretical insights that led to the prediction of new properties.

The inflow of researchers and organisations in graphene research as result of the ‘Novoselov paper’

In some cases discoveries stimulate other researchers to do research related to the discovery, and to publish the outcomes of their research. Our focus is on the diffusion among scientist and research organisations of the knowledge presented in a publication. We question if characteristics of this diffusion process can be used to identify breakthroughs at early stage. To do this we track scientists that become active in graphene research, and that cite the ‘Novoselov paper’ in their publications. To measure this influx of researchers we

Table 2 Citations to the ‘Novoselov paper’ in the first 12, 24, and 36 months after publication, the ranking among all graphene publications from 2004 and 2005, and the ranking among all publications from the same Volume (306) of Science on the basis of the number of citations

Citation window (: time lag between publication and citation) (months)	Number of citations to the ‘Novoselov paper’	Ranking among all graphene publications from 2004 and 2005 (%)	Ranking among all publications of Science Volume 306 (%)
≤12	18	Top 4	Top 30
≤24	105	Top 3	Top 5
≤36	373	Top 1	Top 1

focus on first authors of publications and their affiliations.²² Authors and affiliations are counted only the first time they publish a publication citing the ‘Novoselov paper’; we consider this moment in time the moment the author and his or her organisation enters graphene research stimulated by the ‘Novoselov paper’.

To measure this influx of authors and organisations we define two quantities *New First Authors (NFA)* and *New First main Organisations (NFO)*. These quantities measure the number of first authors and their affiliations that become active within a time interval, and who have not been mentioned before as first author or first organisation on a publication citing the ‘Novoselov paper’. The set of distinct authors at time T is presented as $N_{A,T}$, and the set of distinct main organisations at time T as $N_{O,T}$. For a period ranging from t_1 until t_2 *NFA* is defined as $N_{A,t_2} - N_{A,t_1}$, and *NFO* as $N_{O,t_2} - N_{O,t_1}$. Based on these two quantities we define the relative mutation of *NFA* (*RelMutNFA*) as $\frac{N_{A,t_2} - N_{A,t_1}}{N_{A,t_1}}$, and *NFO* (*RelMutNFO*) as $\frac{N_{O,t_2} - N_{O,t_1}}{N_{O,t_1}}$. These two quantities measure *NFA* and *NFO* in relation to the number of distinct authors and distinct organisations that are already active at the beginning of the interval (t_1). The ratio of *NFA* and *NFO* (*NFANFORatio*) indicates if authors and organisations become active at the same pace, and is defined as $\frac{N_{A,t_2} - N_{A,t_1}}{N_{O,t_2} - N_{O,t_1}}$. We define two more quantities *NFAcitsratio* $\left(\frac{Ncits_{t_2} - Ncits_{t_1}}{N_{A,t_2} - N_{A,t_1}}\right)$, and *NFOcitsratio* $\left(\frac{Ncits_{t_2} - Ncits_{t_1}}{N_{O,t_2} - N_{O,t_1}}\right)$. These quantities compare the number of authors, and organisations that become active with the increase of the citation count in the same period. If the increase of the citation count is the result solely of authors or organisations entering the field the ratio is close to 1.0. When already active authors and organisations continue producing citations the ratio is above 1.0.

We compared the values for publications²³ citing the ‘Novoselov paper’ with the values for 318 publications from Volume 306 of Science, and the 66,073 publications from October 2004—the month of the ‘Novoselov paper’ was published—to find out if the proposed quantities have a discriminating effect on the publications. In all cases citations within 24 months after the publication of a cited document are taken into account. First for each cited publication the values were calculated separately. The distributions of the values for the various quantities were computed from the values obtained for the individual publications. Table 4 shows that for the ‘Novoselov paper’ *NFA*, *NFO*, *RelMutNFA*, and *RelMutNFO* are in the top 5 % and in a number of cases in the top 1 %. The *NFANFORatio*

²² First author = the author first mentioned on a publication, first organisation = the affiliation of the first author of a publication. Ambiguity in names and the effects for the analysis are discussed in section ‘Influence of ambiguity of author and organisation names’.

²³ Only publications covered in TR/CWTS WoS of type article and letter are taken into account to focus on publications containing original research.

Table 3 Journal classification science levels

Category number	Science level
1	Discovery science
2	Industrially relevant science
3	Science-based technological development
4	Clinically relevant science
5	Science-based clinical practice
6	Industrial/medical development

for the ‘Novoselov paper’ is in the top 50 % for the publications in Science, and in the top 10 % of the publications from October 2004. *NFAcitsratio* for the ‘Novoselov paper’ is in the top 5 % of the publications in Science Volume 306, and in the top 11 % for the publications from October 2004. *NFOcitratio* is in the top 25 % the publications in Science Volume 306, and in the top 13 % for the publications of October 2004.

Influence of ambiguity of author and organisation names

The names of authors and organisation in the Web of Science database are known to be ambiguous in several cases. This ambiguity complicates the identification of unique authors and unique main organisations. In the computations we used the names as they appear in the database (TR/CWTS WOS) without any further disambiguation. To check for the effects of ambiguous names we took a random sample of 1,000 publications out of the 66,073 publications from October 2004. For each publication in this sample only the publications citing it within 36 months were considered; the first authors names and their affiliations were collected, and checked for obvious data errors. The author names did not reveal any obvious errors. In 18 cases the main organisation name was not available in the database, and in 10 cases the main organisation names contained obvious errors such as misspellings or inconsistent use of abbreviations. These latter errors could be corrected easily. A further check was done on the name of the first author in combination with the main organisation name. If a combination of author name and organisation name was identical for publications citing the same document the names were considered to be unambiguous for the cited publication. The occurrence of multiple organisations for the same first author citing the same publication was found in 341 cases. In these cases more than one author might be involved; other options such as authors with multiple affiliations, and authors switching from on organisation to another during the period are possible. We conclude that in 341 (9 %) cases the author names might be ambiguous. No cases were found for which the group of articles and letters citing the same publication contained an ambiguous main organisation name.

Expert opinion

We discussed preliminary findings of this study with Frenken (2013), a senior scientist active in graphene research. Table 5 briefly presents information on Frenken’s professional background.

The discussion focused on three questions.

1. *Did the ‘Novoselov paper’ introduce new theoretical concepts, a paradigm shift?*
As motioned in (Nobel Prize Physics, 2010) Wallace (1947) and Boehm et al. (1962)

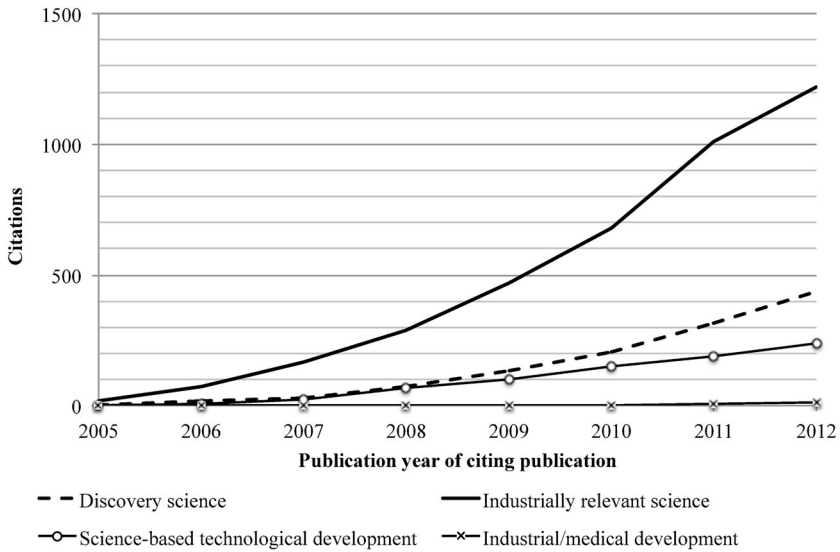


Fig. 3 Scholarly publications citing the ‘Novoselov paper’ disaggregated into science levels

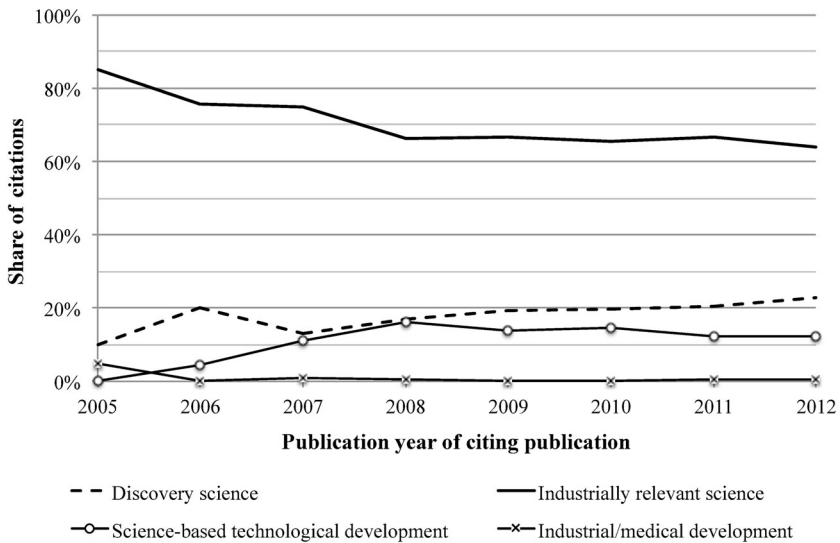


Fig. 4 Publications per science category as share of the total number of scholarly publications citing the ‘Novoselov paper’

presented the theoretical background for the graphene breakthrough published in 2004. The discovery in the ‘Novoselov paper’ is about producing, isolating, identifying and characterizing graphene and as such does not introduce a paradigm shift.

2. *What causes the dichotomy between applied science and theoretical science that becomes apparent in 2007–2008 in the co-occurrence maps of terms from the*

abstracts and the titles of graphene publications, and that evolves into a (meta) stable configuration?

The observed dichotomy signals the renewed interest from theoretical physicists for graphene leading to the application of much more modern theories than those available to Wallace in 1947 or Boehm in 1962. The moment experiments showed that the predicted properties for graphene existed in reality, theoretical physicist became again interested and started to apply more recent theoretical approaches. These new approaches led to the prediction of new properties. The consequence is that publications focused on theoretical aspects of graphene kept appearing and together with the up rise of publications on applied graphene research led to the division of graphene publications in two main areas.

3. *What is the challenge in contemporary graphene research?*

The current challenge is to produce sheets of graphene that are substantially bigger than the small pieces, roughly 1 cm^2 , that can be produced at this moment.

Discussion and concluding remarks

The goal of this study to prove our hypothesis ‘Bibliographic data contains information that enables to identify and characterize at early stage publications of potential breakthroughs at the interface of science and technology’ to be true. We analysed bibliographic data for the publication by Novoselov et al. from 2004 as a well-known example of a discovery that is generally considered a breakthrough. In our research we differentiate breakthroughs using Koshland’s Cha–Cha–Cha-theory. We do this as we think that breakthroughs leave behind patterns in bibliographic data that reflect the characteristics that are typical for the type of the breakthrough. This study concentrates on five statements to uncover bibliographic information that could help in identifying at early stage publications containing a potential breakthrough discovery.

1. *The scientific discovery has the potency to evolve into a new technology discovery.* We view patent publications as representatives of technological developments. Scholarly publications cited in patents, especially those cited by patent applicants, show direct links between science and technology. The fact that the ‘Novoselov paper’ is cited in patents shows that this publication links science and technology. Figure 1 shows furthermore that from 2006 onwards the number of ‘graphene’ related patent filings increased following with a time lag the uprise in the number of graphene related scholarly publications, and especially the publications citing the ‘Novoselov paper’.
2. *If the discovery does introduce a new theoretical framework it is classified as a ‘challenge’ discovery, otherwise it is classified as a ‘charge’ discovery.* The subject of the ‘Novoselov paper’ is obtaining and identifying graphene, and it is therefore a ‘technical’ publication providing a method how to obtain this almost mythical material to do experiments on. Our discussion with professor Frenken confirmed that the ‘Novoselov paper’ did not introduce a new theoretical framework. By assigning *science levels* to the publications citing the ‘Novoselov paper’ and analysing the evolution of the number of publications assigned to these science levels supports this view. This characterizes the discovery in the ‘Novoselov paper’ as a charge breakthrough as breakthroughs of this type, by definition, do not introduce paradigm shifts.

Table 4 The quantities *NFA*, *NFO*, *RelMutNFA*, *RelMutNFO*, *NFANFOratio*, *NFAcitsratio*, *NFOcitsratio* for articles and letters citing the ‘Novoselov paper’ compared with the values obtained for publications the 318 publications from Science Volume 306, and the 66,073 publications published in October 2004

Quantity	Publication(s)		Percentile borders							
			0 %	25 %	50 %	75 %	90 %	95 %	99 %	100 %
<i>NFA</i>	‘Novoselov paper’	52.0								
	Science Volume (306)		0.0	3.0	0.5	19.0	30.0	41.7	69.0	88.0
	October 2004		0.0	0.0	1.0	3.0	5.0	7.0	15.0	163.0
<i>NFO</i>	‘Novoselov paper’	41.0								
	Science Volume (306)		0.0	3.0	8.0	14.0	21.3	29.7	48.0	58.0
	October 2004		0.0	0.0	1.0	2.0	5.0	6.0	13.0	100.0
<i>RelMutNFA</i>	‘Novoselov paper’	5.8								
	Science Volume (306)		0.0	1.0	1.7	2.4	4.0	5.1	8.1	18.0
	October 2004		0.0	0.0	0.0	1.0	2.5	4.0	7.0	22.0
<i>RelMutNFO</i>	‘Novoselov paper’	4.6								
	Science Volume (306)		0.0	0.8	1.3	2.2	3.4	4.3	9.3	17.0
	October 2004		0.0	0.0	0.0	1.0	2.0	3.0	6.0	22.0
<i>NFANFOratio</i>	‘Novoselov paper’	1.3								
	Science Volume (306)		0.0	1.0	1.2	1.4	1.7	1.8	2.2	3.0
	October 2004		0.0	0.0	1.0	1.0	1.3	1.6	2.3	7.0
<i>NFAcitsratio</i>	‘Novoselov paper’	1.4								
	Science Volume (306)		1.0	1.0	1.0	1.2	1.3	1.4	1.7	2.0
	October 2004		1.0	1.0	1.0	1.0	1.5	2.0	3.0	11.0
<i>NFOcitsratio</i>	‘Novoselov paper’	1.7								
	Science Volume (306)		1.0	1.1	1.3	1.7	2.0	2.3	3.0	5.0
	October 2004		1.0	1.0	1.0	1.3	2.0	2.0	4.0	16.0

Citations within 24 months after the publication of a cited document are considered

3. *The publication is highly cited from the moment it is published.* As shown in Fig. 2 and Table 2 the ‘Novoselov paper’ from its publication received many citations. It received these citations well before the Nobel Prize Physics 2010 was awarded.
4. *The discovery leads to renewed interest from theoretical physicists.* The evolution of co-occurrence maps (Winnink and Tijssen 2014) based on the terms used in the publications citing the ‘Novoselov paper’ revealed that from 2007 onwards the terms clearly cluster in two areas. Both areas are in ‘balance’; one cluster denotes basic research and the second cluster focuses on applied research. The results presented in Fig. 4 support this conclusion that is also in line with information from Frenken. The conclusion is that theoretical physicists regained interest in graphene around 3 years after the publication of the ‘Novoselov paper’. This renewed interest was not the result of a paradigm shift, but was driven by the outcomes of experimental research that

Table 5 Information on the professional background of professor Frenken

<p>Joost Frenken is a Full Professor at the Physics Department (LJON) of Leiden University, The Netherlands, where he is head of the Interface Physics Group</p> <p>□ Central to the research of Joost Frenken is the fundamental understanding of the dynamic aspects of surfaces and interfaces and their role in relevant, natural or industrial processes under practical conditions. Topics of interest include surface diffusion, crystal growth, surface phase transitions, catalysis, and friction. For tailor-made measurements in each of these areas, Frenken's research group has developed a variety of special-purpose scanning probe microscopes as well as dedicated instrumentation for surface X-ray diffraction</p> <p>□ Frenken has been (co)-initiator of two spin-off companies. Leiden Probe Microscopy BV markets commercialized versions of the research equipment developed in Frenken's research group, with an emphasis on advanced scanning probe and X-ray diffraction instrumentation. Applied NanoLayers BV is a recent spin-off, specialized on the production of high-quality graphene and boron nitride monolayers (http://ip.physics.leidenuniv.nl/index.php/people/frenken/9-cat-members/51-frenken-bio) (2013-12-23)</p> <p>As of January 1st 2014, Joost Frenken will be the director of a new scientific research center. This research facility focuses on physical and chemical processes that are crucial to extreme UV lithography. This new research center, ARCnL, is a collaboration between ASML, FOM, NWO and the two universities in Amsterdam (University of Amsterdam and the VU). (http://ip.physics.leidenuniv.nl) (2013-12-23)</p>

- confirmed predictions already made by Wallace (1947) and Boehm et al. (1962). Theoretical physicists became interested and applied more recent theoretical insights to the ‘graphene issue’ and came up with new predictions (Frenken 2013).
5. *The discovery results in an above average influx of researchers and organisations new to the research field.* Using the quantities *NFA*, *NFO*, *RelMutNFA*, *RelMutNFO*, *NFANFOratio*, *NFAcitsratio*, and *NFOcitsratio* we try to measure the diffusion of the knowledge presented in a particular publication among researchers and research institutes. This study indicates that these quantities have discriminating power, and might be useful to classify publications between those containing a potential breakthrough discovery and those that do not. For most quantities the ‘Novoselov paper’ gets high values. The use of these measures presupposes the unambiguity of names of authors and research organisations; in general this is not the case. We checked the ambiguity of names and found that it plays a minor role. The reason for this is that the probability for a name occurring more than once on publications that cite the same document in a short period, 36 months or less, is limited. During testing we found that 9 % of the names might be ambiguous.

In this study we used Novoselov et al. (2004) as typical example of a breakthrough at the interface of science and technology. Citation relations form the basis in our analysis. The analyses of the effects of ‘Novoselov paper’ on graphene R&D shows that it is a typical example of a ‘charge’ breakthrough, and that no paradigm shift is involved. We propose a set of measures to be derived from bibliographic data that could help in identifying at early stage publications that contain a potential scientific breakthrough. The proposed measures have discriminating power. Ambiguity of names showed not to be an issue for these measures due to the short time period used in the analyses in combination with the focus on citations to one particular publication—in this case the ‘Novoselov paper’. The importance of identification of breakthroughs in science at early stage is that it could aid policy makers, funding organisations, and companies to timely prioritize resources for R&D.

The scientific community is, in this study, seen as an analytical instrument that evaluates scientific discoveries, and as result generates 0 or more citing publications. The bibliographic information for these citing publications reflects the opinion of the scientific community on the cited publication. The measures we derived in this study suggest that they can be used to differentiate publications into those that probably contain a potential breakthrough and those not. We do realize that this case study is an analysis of only one well-known breakthrough; the measures we propose need to be validated for general applicability; such a study is foreseen as a follow up. In this follow up study we will test the outcomes of this and other studies on random sets of publications.

Our intention with this study was to find information in bibliographic data that might be the basis for indicators that could help in identifying and characterizing at early stage potential breakthrough discoveries at the interface of science and technology. Not addressed in this study is the fact that discoveries considered at first a breakthrough might at a later stage prove not to be. The importance of identification of breakthroughs in science at early stage is that it could aid policy makers, funding organisations, and companies to timely prioritize resources for R&D.

Acknowledgments The authors kindly thank Prof. J.W.M. Frenken, Leiden Institute of Physics—Leiden University, and Prof. A.F.J. van Raan, Centre for Science and Technology Studies—Leiden University, for their useful comments on an earlier version of this manuscript. We also like to thank the anonymous reviewers for their valuable comments during peer review of this manuscript.

References

- Adams, J. (2005). Early citation counts correlate with accumulated impact. *Scientometrics*, 63, 567–581. doi:10.1007/s11192-005-0228-9.
- Andersen, P. D., & Borup, M. (2009). Foresight and strategy in national research councils and research programmes. *Technology Analysis & Strategic Management*, 21(8), 917–932.
- Bañuls, V. A., & Salmeron, J. L. (2008). Foresighting key areas in the information technology industry. *Technovation*, 28(3), 103–111.
- Bettencourt, L. M., Kaiser, D., Kaur, J., Castillo-Chávez, C., & Wojcik, D. (2008). Population modeling of the emergence and development of scientific fields. *Scientometrics*, 75(3), 495–518.
- Boehm, H., Clauss, A., Hofmann, U., & Fischer, G. (1962). Dunnste kohlenstofffolien. *Zeitschrift für Naturforschung Part B-Chemie Biochemie Biophysik Biologie und Verwandten Gebiete, B*, 17(3), 150–153.
- Breiner, S., Cuhls, K., & Grupp, H. (1994). Technology foresight using a Delphi approach—A Japanese–German cooperation. *R&D Management*, 24(2), 141–153.
- Chen, C., Chen, Y., Horowitz, M., Hou, H., Liu, Z., & Pellegrino, D. (2009). Towards an explanatory and computational theory of scientific discovery. *Journal of Informetrics*, 3(3), 191–209.
- Coates, V., Farooque, M., Klavans, R., Lapid, K., Linstone, H. A., Pistorius, C., et al. (2001). On the future of technological forecasting. *Technological Forecasting and Social Change*, 67(1), 1–17.
- Coates, J. F., Mahaffie, J. B., & Hines, A. (1994). Technological forecasting: 1970–1993. *Technological Forecasting and Social Change*, 47(1), 23–33.
- Daim, T. U., Rueda, G., Martin, H., & Gerdri, P. (2006). Forecasting emerging technologies: Use of bibliometrics and patent analysis. *Technological Forecasting and Social Change*, 73(8):981–1012. Tech Mining: Exploiting Science and Technology Information Resources.
- Frenken, J. W. M. (2013). Personal communication (unpublished).
- Geim, A. K., & Novoselov, K. S. (2007). The rise of graphene. *Nature Materials*, 6(3), 183–191.
- Grupp, H., & Schmoch, U. (1992). *Dynamics of science-based innovation, chapter 9—At the crossroads in laser medicine and polyimide chemistry: Patent assessment of the expansion of knowledge*, (pp. 269–301). Berlin: Springer.
- Hand, D. J. (2009). Mining the past to determine the future: Problems and possibilities. *International Journal of Forecasting*, 25(3):441–451. Special Section: Time Series Monitoring.
- Hollingsworth, J. R. (2008). Scientific discoveries: An institutionalist and path dependent perspective. In Hannaway, C. (Ed.), *Biomedicine in the twentieth century: Practices, policies, and politics* (pp. 317–353). Amsterdam: IOS Press.
- Isenson, R. S. (Ed.). (1969). Project hindsight (final report). Technical report, US Dept. of Defense.
- Jewkes, J., Sawers, D., & Stillerman, R. (1958, 1969). *The sources of invention*. London: MacMillan.
- Julius, M., Berkoff, C. E., Strack, A. E., Krasovec, F., & Bender, A. D. (1977). A very early warning system for the rapid identification and transfer of new technology. *Journal of the American Society for Information Science*, 28(3), 170–174.
- Koshland, D. E. (2007). The Cha-Cha-Cha theory of scientific discovery. *Science*, 317(5839), 761–762.
- Kuhn, T. S. (1962). *The structure of scientific revolutions*. Chicago: The University of Chicago Press.
- Leydesdorff, L., & Rafols, I. (2011). 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.
- Martin, B. R. (1995). Foresight in science and technology. *Technology Analysis and Strategic Management*, 7(2), 139–168.
- Meade, N., & Islam, T. (1998). Technological forecasting—Model selection, model stability, and combining models. *Management Science*, 44(8), 1115–1130.
- Mishra, S., Deshmukh, S. G., & Vrat, P. (2002). Matching of technological forecasting technique to a technology. *Technological Forecasting and Social Change*, 69(1), 1–27.
- Nobel Prize Physics. (2010). *Scientific background on the Nobel Prize in physics 2010—GRAPHENE, compiled by the class for physics of the Royal Swedish Academy of Sciences*. http://www.nobelprize.org/nobel_prizes/physics/laureates/2010/advanced-physicsprize2010.pdf.
- Novoselov, K., Geim, A., Morozov, S., Jiang, D., Zhang, Y., Dubonos, S., et al. (2004). Electric field effect in atomically thin carbon films. *Science*, 306(5696), 666–669.
- Ponomarev, I. V., Williams, D. E., Hackett, C. J., Schnell, J. D., & Haak, L. L. (2014). Predicting highly cited papers: A method for early detection of candidate breakthroughs. *Technological Forecasting and Social Change*, 81, 49–55.
- Ponomarev, I., Williams, D., Lawton, B., Cross, D. H., Seger, Y., Schnell, J., et al. (2012). Breakthrough paper indicator: Early detection and measurement of ground-breaking research. In Jeffery, K. G., &

- Dvořák, J. (Eds.), *E-Infrastructures for research and innovation: Linking information systems to improve scientific knowledge production: Proceedings of the 11th international conference on current research information systems* (pp. 295–304).
- Shen, Y. C., Chang, S. H., Lin, G. T., Yu, H. C., et al. (2010). A hybridselection model for emerging technology. *Technological Forecasting and Social Change*, 77(1), 151–166.
- Small, H. (1977). Co-citation model of a scientific specialty—Longitudinal-study of collagen research. *Social Studies of Science*, 7(2), 139–166.
- Tijssen, R. J. W. (2010). Discarding the ‘basic science/applied science’ dichotomy: A knowledge utilization triangle classification system of research journals. *Journal of the American Society for Information Science and Technology*, 61(9), 1842–1852.
- Tu, Y.-N., & Seng, J.-L. (2012). Indices of novelty for emerging topic detection. *Information Processing and Management*, 48(2), 303–325.
- Van Andel, P. (1994). Anatomy of the unsought finding. Serendipity: Origin, history, domains, traditions, appearances, patterns and programmability. *The British Journal for the Philosophy of Science*, 45(2), 631–648.
- Van Raan, A. F. (2004). Sleeping beauties in science. *Scientometrics*, 59(3), 467–472.
- Wallace, P. R. (1947). 1946 Annual meeting at New York, January 30, 31 and February 1, 1947. *Physical Review*, 71, 460–489.
- Winnink, J. J. (2012). Searching for structural shifts in science: Graphene R&D before and after Novoselov et al. (2004). In Archambault, E., Gingras, Y., & Larivière, V., (Eds.), *Proceedings of the 17th international conference on science and technology indicators (STI 2012)* (Vol. 2, pp. 837–846).
- Winnink, J., & Tijssen, R. (2014). Early stage identification of ‘charge’ breakthroughs at the interface of science and technology: Lessons drawn from a landmark publication. *Supplementary material*.
- Yoon, B., & Park, Y. (2007). Development of new technology forecasting algorithm: Hybrid approach for morphology analysis and conjoint analysis of patent information. *IEEE Transactions on Engineering Management*, 54(3), 588–599.
- Young, P. (1993). Technological growth curves: A competition of forecasting models. *Technological Forecasting and Social Change*, 44(4), 375–389.