Analysing the science-technology link using interlinked databases

Jos	Winnink	winninkjj@cwts.leidenuniv.nl	Netherlands	CWTS - Leiden University

Introduction

Several scholars tried to unravel the processes behind the evolution of science, as advances in science are seen the driving force behind technological developments that have a major impact on the economy and society. It is for this reason that many governments focus on policy measures or policy programs to stimulate technological innovation (OECD, 1992) as technological change is commonly considered the main determinant of economic growth. Seminal work in the 1950s and 1960s (Smits, 2002) is considered the main driving force between the analyses of the science-technology interaction. Jewkes et al. (1958) is one of these early seminal systematic studies; two other studies - the ‘Hindsight project’ (Isenson, 1969) and the ‘Batelle study’ (Globe et al., 1973) - are also well recognized for their comprehensive and systematic analysis of the interaction between science and technology. All these studies have in common that they analysed, sometimes large numbers of, individual cases, without the availability of large (bibliographic) databases with relevant information that could be analysed in a systematic way using computers. Martin (1995) concludes ‘… in using foresight to help in selecting and exploiting research that is likely to yield longer-term economic and social benefits …’.

It is well known that not all scientific discoveries in science have the same impact on future developments of science. Some scientific discoveries have a major impact – the ‘breakthroughs’. In as similar way technological inventions exist that are considered ‘breakthrough inventions’ as they have a major iimpact on the development of a particular technology. Methods to identify scientific discoveries that are potential breakthroughs and also methods to identify breakthrough inventions are subject of research by several scholars. Identifying potential breakthrough papers are proposed for instance by Redner (2005), Schneider and Costas (2015), Ponomarev et al. (2014) and Winnink (2017). Some of these methods focus on automatic computerised methods that facilitate early stage identification of such papers. Squicciarini et al. (2013) provides a definition that can be used to identify breakthrough inventions.

General interest exists in the factors that govern the mutual influences between science and technology. Bettencourt et al. (2009) analysed the inception and evolution of eight scientific fields, and show that a number of universal features govern the evolution of a scientific field. The availability of databases with bibliographic information on patents and bibliographic information on scientific publications enables the automatic computerized large-scale exploration of links between science and technology. Such explorations can reveal mechanisms of the way scientific discoveries and the development of new technologies are interrelated.

Methodology

We implemented a data infrastructure to analyse relations between science and technology. This infrastructure is composed of three major components: (1) a licenced in-house version of Clarivate Analytics’ Web of Science database (WoS) containing information on scientific papers published form 1980 onwards; (2) the PATSTAT database provided by the European Patent Office (EPO) containing bibliographic information of more than 90 million patent documents; and (3) an in-house developed database that links scientific publications cited in patents to corresponding records in the WoS.

The computerised algorithms developed and implemented to identify breakthrough publications (Winnink, 2017) are applied to the data in the WoS to generate a dataset that consists exclusively of potential breakthrough discoveries in science. Furthermore we applied the OECD-definition for ‘breakthrough inventions’ (Squicciarini et al., 2013) to construct a dataset of ‘breakthrough inventions’. We use bibliographic information from the WoS as a proxy for the development of science and use patent information from PATSTAT as a proxy for technological developments. Using these information sources we are able to analyse not only the relation between scientific discoveries and technological developments in general, but are also able to analyse how breakthrough discoveries and breakthrough inventions interact.

A preliminary result and future research

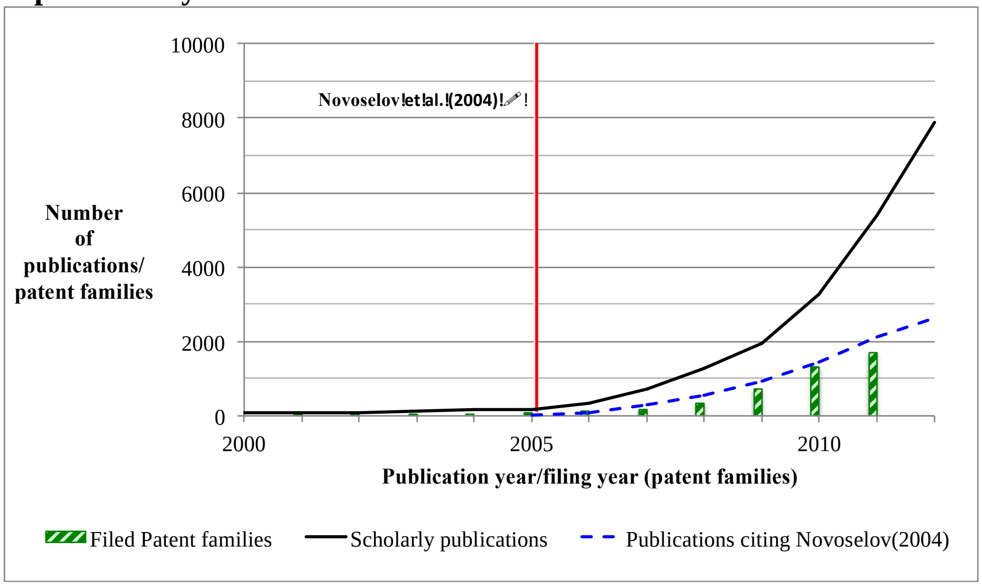

Figure 1 Development of the Graphene field (scientific publications, patent publications)

In Figure 1 the evolution of graphene-related research is shown by the number of graphene related publications in the WoS and the number of patent applications covered by the PATSTAT database. The moment the breakthrough publication on 'graphene' (Novoselov et al., 20014), for which the Nobel Prize Physics 2010 was awarded to Novoselov and Geim, is clearly indicated. The graph shows (blue dashed line) the impact of the paper on subsequent scientific papers.

But what was its direct impact on the development of graphene-related technology? Novoselov et al. (2004) was cited 174 times in patents applied for in the period 2004-2013; 31 of these citing patents are seen as breakthrough inventions according to de definition in Squicciarini et al. (2013) . Some of the questions we focus on in our current research are:

- What is the time lapse between the publication of a scientific paper and the moment it is cited for the first time in a patent?
- Are 'potential breakthrough' papers the main sources for new technologies?
- Are all 'breakthrough patents' science based?
- Can we detect science-technology feedback loops?
- …

More and more detailed results will be presented at the conference.

References

Bettencourt, LMA, Kaiser, DI, and Kaur, J. (2009). Scientific discovery and topological transitions in collaboration networks, *Journal of Informetrics.* 3(3):210–221. Science of Science: Conceptualizations and Models of Science.

Globe, S., Levy, G. W., and Schwartz, C. M. (1973). Science, technology and innovation. NSF-study NSF-C-667, Battelle Memorial Inst., Columbus, Ohio, Columbus Labs.

Isenson, R. S. (1969). Project hindsight (final report). Technical Report AD495905, Office of the Director of Defense Research Engineering, Washington, DC, 20301.

Jewkes et al., 1958 Jewkes, J., Sawers, D., and Stillerman, R. (1958). The Sources of Invention. MacMillan, London.

Martin, B. (1995). Foresight in science and technology, *Technology Analysis & Strategic Management*. 7(2):139–168.

Novoselov, K. S., Geim, A. K., Morozov, S. V., Jiang, D., Zhang, Y., Dubonos, S. V., Grigorieva, I. V., and Firsov, A. A. (2004). Electric field effect in atomically thin carbon films. *Science*, 306(5696):666–9.

OECD (1992). Technology and the economy: the key relationships, volume 42 of TEP - The Technology / Economy Programme. Organisation for Economic Co-operation and Development

Ponomarev, IV, Williams, DE, Hackett, CJ, Schnell, JD, and Haak, LL. (2014). Predicting highly cited papers: A method for early detection of candidate breakthroughs. *Technological Forecasting and Social Change.* 81(January 2014):49–55.

Redner. S. (2005). Citation statistics from 110 years of physical review, *Physics Today*. 58(6):49–54.

Schneider, JW and Costas, R. (2015). Identifying potential breakthrough publications using refined citation analyses: Three related explorative approaches. Available as arXiv:1512.01388v1. pages 1–33.

Smits, R. (2002). Innovation studies in the 21st century: Questions from a user's perspective. *Technological Forecasting and Social Change.* 69(9):861–883.

Squicciarini, M., H. Dernis and C. Criscuolo (2013), "Measuring Patent Quality: Indicators of Technological and Economic Value", OECD Science, Technology and Industry Working Papers, 2013/03, OECD Publishing. http://dx.doi.org/10.1787/5k4522wkw1r8-en

Winnink, J. J. (2017). Early-stage detection of breakthrough-class scientific research: using micro-level citation dynamics. PhD thesis, Social and Behavioural Sciences, Leiden University.