Finding the factors behind potential breakthrough papers

Introduction

As technological change is commonly considered the main determinant of economic growth since seminal work in the 1950s and 1960s (Smits, 2002) many governments focus on policy measures or policy programs to stimulate technological innovation (OECD, 1992). 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. Martin (1995) concludes '… in using foresight to help in selecting and exploiting research that is likely to yield longer-term economic and social benefits …'. 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. A general interest exists therefore in identifying papers that present potential breakthroughs, and especially in factors that cause such papers to have a major impact on developments in science and technology.

Methods to identify potential breakthrough papers are proposed for instance by Redner (2005), Schneider and Costas (2015), and Ponomarev et al. (2014). Winnink et al. (2016) focus on automatic computerised algorithms that facilitate the early stage identification of such papers. These algorithms harvest at large scale bibliographic information. As expert opinions and time are needed to cast a judgement if the research findings in the paper represent a broadly accepted scientific breakthrough we call these potential breakthrough papers 'breakout' papers. The early stage detection algorithms were adapted so they are not only suitable for the early stage identification but can also be used to analyse the breakout character of individual papers at any point in time. These adapted algorithms allow us to analyse the dynamic character of 'environmental' factors that cause a paper to become a breakout paper. Our primary focus is on the collaboration of authors. The factors we currently investigate are (1) the type of the organisations with which collaborating authors are affiliated, (2) cross border collaboration –local, domestic, and international-, and (3) the size of the research group. The first moment a paper is identified as a 'breakout' paper by one of the algorithms is used as point of reference. Special attention is given to the question 'Does the influence of these factors depend on the age of a paper?' This analysis should also help in unravelling the factors that influence delayed recognition of a paper, including the mysteries behind the awakening of sleeping beauties (van Raan, 2004).

Methodology

The computerised algorithms that we developed and implemented (Winnink et al., 2016) are applied to the data in the CWTS-licenced in-house version of Thomson Reuters Web of Science database (WoS). From this database we selected all scholarly papers of WoS-type 'article' and 'letter' that are published between 1990 and 1994. Applying the algorithms results in a set of papers of which each one is classified as a breakout paper by at least one of the algorithms. We left out papers from two of the seven NOWT[1]-categories. These categories 'Language, Information and Communication', and 'Law, Arts and Humanities' resulted in to few breakout papers for meaningful analysis.

Preliminary results

Preliminary research shows that 4.3% of the papers is marked as a breakout paper. The majority (98.2%) of the papers get their breakout 'status' within two year after publication. After these two years the number of papers showing breakout characteristics steadily drops as time proceeds. Papers written by authors affiliated with a combination of organisations have a higher probability to be a breakout paper; papers from authors exclusively affiliated with companies (C) or hospitals (H) have a below average

[1] Netherlands Observatory of Science and Technology (NOWT)

chance of being a breakout paper. Table 1 presents some of the preliminary results. More and more detailed results will be presented at the conference.

Table 1 Moment of occurrence of the breakout-character of a paper, 'article' or 'letter', published in the period 1990-1994 vs. organisational collaboration

	Total	**Years since paper a paper is identified as a breakout paper**									
Affiliation type		0	1	2	3	4	5	6	7	8	9
		Number of papers									
University (U)	1886048	79144	4699	362	238	139	209	166	175	118	98
Research Organisation (R)	225731	10046	700	42	20	16	23	19	15	21	13
Company (C)	129950	4557	454	29	18	10	12	14	17	13	6
Hospital (H)	99178	1936	121	8	8	5	9	1	1	3	4
U + R	147424	9691	912	30	31	15	20	28	22	12	10
U + C	76446	5364	575	30	19	8	15	13	10	12	6
U + H	86069	3416	398	31	16	2	8	7	9	11	7
U + H + R	6067	375	99	4	1	0	0	2	2	0	0
U + H + C	2911	200	44	1	3	1	1	0	0	0	0
U + H + C + R	476	49	9	0	0	0	0	0	0	0	0
Total papers	2660300	114778	8011	537	354	196	297	250	251	190	144
		91,8%	*6,4%*	*0,4%*	*0,3%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,2%*	*0,1%*

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.

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

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.

van Raan, AFJ. (2004). Sleeping beauties in science. *Scientometrics*, 59(3):467–472.

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.

Winnink, JJ. Tijssen, RJW, and van Raan, AFJ. (2016). Can early-detection algorithms of breakout papers uncover scientific breakthroughs? *(Submitted for publication)*