

Citation algorithms for identifying research milestones driving biomedical innovation

Jordan A. Cominsa ^{1,*}, Loet Leydesdorff ²

¹ Center for Applied Information Science, Virginia Tech Applied Research Corporation, Arlington, VA, United States.

² Amsterdam School of Communication Research (ASCoR), University of Amsterdam, PO Box 15793, 1001 NG Amsterdam, The Netherlands.

* Corresponding Author. E-mail: jcomins@gmail.com

Abstract. Scientific activity plays a major role in innovation for biomedicine and healthcare. For instance, fundamental research on disease pathologies and mechanisms can generate potential targets for drug therapy. This co--evolution is punctuated by papers which provide new perspectives and open new domains. Despite the relationship between scientific discovery and biomedical advancement, identifying these research milestones that truly impact biomedical innovation can be difficult and is largely based solely on the opinions of subject matter experts. Here, we consider whether a new class of citation algorithms that identify seminal scientific works in a field, Reference Publication Year Spectroscopy (RPYS) and multi--RPYS, can identify the connections between innovation (e.g., therapeutic treatments) and the foundational research underlying them. Specifically, we assess whether the results of these analytic techniques converge with expert opinions on research milestones driving biomedical innovation in the treatment of Basal Cell Carcinoma. Our results show that these algorithms successfully identify the majority of milestone papers detailed by experts (Wong and Dlugosz 2014) – thereby validating the power of these algorithms to converge on independent opinions of seminal scientific works derived by subject matter experts. These advances offer an opportunity to identify scientific activities enabling innovation in biomedicine.

Keywords: Reference Publication Year Spectroscopy; Citation Analysis; Algorithmic Historiography; Bibliometrics.

1. Introduction

Biomedical innovation is guided, at least in part, by historical influences (Agarwal and Searls, 2009). In some cases, biomedical innovation is guided by key scientific discoveries, referred to here as research milestones treatments (Wong and Dlugosz, 2014). Nelson, Buterbaugh, Perl, & Gelijns (2011) distinguish analytically among three enabling forces in medical innovation: advances of scientific understanding of diseases, learning in clinical practices, and advances in technological capabilities (which often originate outside of medicine) for the development of novel modalities of diagnosis and treatment. However, interactions between supply--side factors and demand originating from diseases can be expected to shape co--evolutions along trajectories enabled by technological capabilities (Petersen, Rotolo, & Leydesdorff, 2016). From the historiographic perspective in innovation studies, however, the innovation itself is focal (Von Hippel, 1988); but the scientific knowledge base of innovations can be excavated from the literature (Garfield, Sher, & Torpie, 1964).

Medical innovations are sometimes triggered by breakthroughs and new developments on the supply side (Leydesdorff & Rafols, 2011; Mina, Ramlogan, Tampubolon, & Metcalfe, 2007).

Connecting major advances in medical treatments with earlier research milestones is considered essential for public appreciation of the role of basic research discoveries in major health advances (Williams et al., 2015) and offers an opportunity to more fully consider the factors leading to new drugs (Nelson, Buterbaugh, Perl, & Gelijns, 2011; Swanson, 1990).

Despite the value of the translation of scientific activities into medical innovations (Mogoutov, Cambrosio, Keating, & Mustar, 2008), current approaches for retrieving research milestones require considerable time and commitment of subject matter experts. In this brief communication, we assess

the utility of using computational approaches to identify research milestones, which may allow for broader engagement of scientific and non-scientific communities with a field's intellectual history.

Eugene Garfield, who first created the Science Citation Index, found the development of citation algorithms to reconstruct the intellectual history of scientific fields a top priority for scientometrics (e.g., Garfield et al., 1964). Citations serve as a functional linkage between ongoing scientific efforts with prior endeavors (Price, 1965; Garfield, Malin, & Small, 1978; Radicchi, Fortunato, & Castellano, 2008). The accumulation of citations is therefore commonly used as a proxy for quantifying the scholarly impact of research (Kostoff and Shlesinger 2005; Marx et al. 2014; van Raan 2000).

Recent years have witnessed noteworthy advances in citation algorithms for identifying the seminal works within the history of a research field. In particular, the development of a new technique, known as Reference Publication Year Spectroscopy (RPYS), offers the possibility of visualizing key seminal papers for a research field using publicly available analysis tools (Comins and Hussey 2015a; Comins and Leydesdorff 2016a; Leydesdorff et al. 2014; Marx *et al.* 2014; Thor et al. 2016).¹ Up until now, use of RPYS for identifying seminal works has been limited to scientific fields where the authors can consider themselves as experts (Comins and Hussey 2015a, 2015b; Comins and Leydesdorff 2016a; Elango et al. 2016; Marx and Bornmann 2013; Marx et al. 2014; Wray and Bornmann 2015) and can therefore support the interpretation of results. Ultimately, this means that testing and evaluation of RPYS-based algorithms suffers from a lack of objectivity as the authors themselves also served as “referees” of the quality of RPYS results. Despite the potential of this technique for identifying seminal works, no studies on RPYS to date have compared its performance to the historical account of research milestones developed independently by experts.

This study represents a first attempt to evaluate RPYS performance more objectively by comparing the results of the technique to seminal articles already identified within a research area by subject matter experts. Here, we utilize a recent article published in the Journal of Investigative Dermatology to compare the milestones identified by subject matter experts on basal cell carcinoma (Wong and Dlugosz 2014) to those identified via RPYS analyses.

Demonstrating that RPYS analyses converge on a similar set of milestones as those identified by experts offers the possibility that such tools can support a broader range of needs in science, including rapid identification of truly landmark findings in a field and informing students and non-academics about pivotal discoveries in the history of research areas.

2. Methods

To assess the performance of RPYS in identifying key milestones described by Wong and Dlugosz (2014), we began by devising search queries capturing the main topics discussed by their article. Search strings were generated by using terms found in the title of the paper as well as within the main text. Our goal was to make a series of searches that progressively become more complex – this would enable us to assess the effectiveness of the algorithms in identifying milestones for straightforward searches as well as more sophisticated search strings. Our search strings are shown in Table 1.

Table 1. Web of Science search queries used to identify records related to basal cell carcinoma, the hedgehog signaling pathway and potential treatments.

Web of Science Search	Records	References
A. TS=(basal cell carcinoma* hedgehog)	1148	23,644
B. TS=(hedgehog inhibit* treatment* basal cell carcinoma*)	244	6,920
C. TS=((cyclopia hedgehog) OR (cyclopamine hedgehog))	918	20,131
D. TS=(basal cell carcinoma* smoothened antagonist)	92	2,700

For instance, search A was a topic search for “basal cell carcinoma* hedgehog”, and was the broadest in terms of number of records retrieved and the number of references they cite. These terms

¹ Implementations of RPYS can be retrieved at <http://www.crexplorer.net/> and <http://comins.leydesdorff.net/>. CRExplorer (the first option) offers the additional option of disambiguation of the cited references (Thor et al., 2016), while RPYS i/o (the second option) offers in addition to standard RPYS, Multi-RPYS as a graphically rich visualization (Comins & Leydesdorff, 2016). A pilot version of this same technique (rpys.exe) can be found at <http://leydesdorff.net/software/rpys/> (Marx, Bornmann, Barth, & Leydesdorff, 2014).

were taken directly from the first four words of the title of Wong and Dlugosz (2014), “Basal Cell Carcinoma, Hedgehog Signaling, and Targeted Therapeutics: The Long and Winding Road.” Search B focused the results from A by including terms related to treatment pathway inhibition by performing a topic search for “hedgehog inhibit* treatment* basal cell carcinoma*”. These terms were found in both the title and subtitles of the paper. Search C was a topic search for “(cyclopia and hedgehog) or (cyclopamine and hedgehog)”, thereby covering the relationship between the hedgehog signaling pathway and its relationship to cyclopamine, a teratogen that can cause the birth defect cyclopia. These search terms were extracted from the title of the article, some of the subtitles and some of the main body text. The terms cyclopia and cyclopamine occurred frequently throughout the document (including figures and references), 18 times. Search D was the narrowest, performing a topic search of “basal cell carcinoma* smoothed antagonist”. This search describes the treatment of basal cell carcinoma by the drug Vismodegib, a competitive antagonist of the smoothed receptor, which was mentioned in the article’s main body text and in its figures. For our initial assessment of RPYS methods to identify milestones, the data analyzed represented the union of Searches A--D; for our second assessment, Searches A--D were analyzed individually. Data was downloaded from the Thomson Reuters Web of Science (WoS) in February 2016.

Earlier work provides details on the procedures for performing standard RPYS (Comins and Hussey 2015b; Elango et al. 2016; Leydesdorff et al. 2014; Marx and Bornmann 2013; Marx et al. 2014; Thor et al. 2016; Wray and Bornmann 2015). In brief, the standard RPYS procedure aggregates cited references from a set of citing publications. For each year, the sum of the number of cited references are plotted by their publication year. This data is then de-- trended by taking the difference between the number of cited references for any given reference publication year (*e.g.*, n) from the 5--year median of reference publications ($y-2, y-1, y, y+1, y+2$). Visualizations were produced using procedures similar to those described in Comins & Leydesdorff (2016b). Peak years appear as those where the de--trended value of cited references in a given year is higher than those of neighboring years. Multi--RPYS extends the impact of these techniques by first segmenting the data in terms of the publication years of the citing sets, performing the standard RPYS analysis within each set and then rank transforming the de--trended results to compare influential references across the history of the citing set (Comins & Leydesdorff, 2016a; Comins & Leydesdorff, 2016b).

To determine the performance of RPYS in identifying Wong and Dlugosz’s milestones (2014; figure 2), we defined success by two criteria. First, our RPYS results needed to reveal a peak year corresponding to the years shown in Wong and Dlugosz milestones figure. Second, these peak years must list the key articles described by Wong and Dlugosz as one of the 10 most referenced articles in that year. Use of CRExplorer (available at <http://www.crexplorer.net>) to clean references enhanced our ability to determine whether a seminal milestone paper was indeed amongst the 10 most referenced articles in a given year (Thor et al., 2016). Without disambiguation of cited reference, attributing reference counts to articles could yield noisy results – leveraging CRExplorer to disambiguate cited references helps to mitigate this concern. Finally, visualizations were generated using the online tool RPYS i/o (at <http://comins.leydesdorff.net>), which conducts both standard and multiple RPYS analyses on WOS datasets.

3. Multi--RPYS for identifying years with research milestones

Using four informed searches (Table 1), we retrieved between 92 and 1,148 records at the Web of Science and these records contained between 2,700 to 23,644 cited references.

Across the 2,402 total records retrieved from these four searches, 1,948 were unique records with 33,481 references (after disambiguation of cited references using clustering methods; see Thor et al, 2016). Multi--RPYS analysis of these 1,948 records (shown in Figure 1) visually reveal multiple bands when potentially influential references were published.

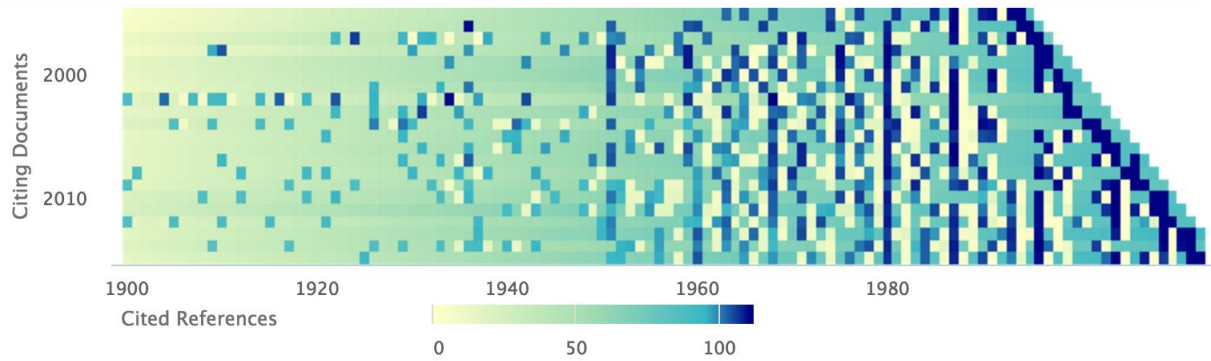


Figure 1. Multi--RPYS results for the union of the four searches outlined in Table 1. Cited References are placed along the x--axis and citing documents along the y--axis. Color represents the cell's ranked RPYS results (see methods) and bands, or columns, in the heatmap are demonstrative of persistently influential works. For example, the band in 1980, is driven by the Nobel prize winning work of Nusslein--Volhard and Wieschaus.

To assess which ones appear most *important*, we submitted results from multi--RPYS to inferential statistical tests to ascertain the cited years that appear most indicative of a milestone year. Confirming the intuition provided by our visualization in Figure 1, we found a significant difference across the cited reference years ($F = 12.94$; $df = 113$; $p < 0.0001$). From here, we assessed which cited reference years appear to deviate the most, indicative of a milestone year. We did so by utilizing a Tukey's HSD post--hoc test and identifying the ten cited reference years with the largest Least Square Means Differences: (1) 1980, (2) 2009, (3) 2012, (4) 1987, (5) 2004, (6) 1996, (7) 2011, (8) 2013, (9) 1968 and (10) 2006.

From this list, 70% matched the milestone years discussed by Wong and Dlugosz (2014). Interestingly, Wong and Dlugosz (2014) designate in total 25 milestones across 18 separate years from 1827 until 2012. Statistically, this would suggest that the chance of correctly identifying a milestone year from 1827--2012 by chance is approximately 10% (18 milestone years / 185 years under study). Our results using multi--RPYS far exceed chance expectations for identifying milestone years and suggests that the combination of multi--RPYS and inferential statistics can rapidly identify milestone years in the history of a research field. In the next section, we more thoroughly examine whether RPYS technique not only identify milestone years, but the precise articles noted by experts as seminal. To do so, we take a methodical approach, using standard RPYS analyses on each of the four individual WoS record searches and describing their associated results.

4. Standard RPYS for identifying specific milestone discoveries

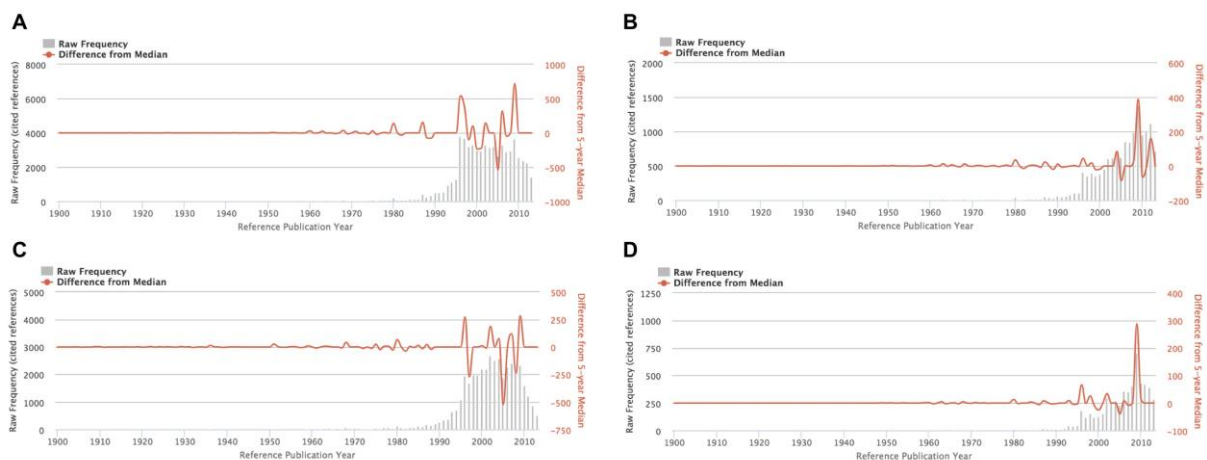


Figure 2. RPYS results for the four searches outline in Table 1. Panels A--D correspond to searches A--D described in Table 1, respectively.

In this section, each set of records from Searches A--D (Table 1) was submitted to its own RPYS analysis (Figure 2) and compared to the seminal papers discussed by Wong and Dlugosz (2014). In

this context, seminal papers were determined by mapping publications from the authors' bibliography to the milestones described in their article (see Wong and Dlugosz, 2014; figure 2). In addition, several milestones were attributed to multiple publications, bringing a grand total of 30 seminal works described by Wong and Dlugosz (2014). Across the RPYS analyses of our four Web of Science searches we successfully capture 23 out of the 30, or 76%, of these milestone articles.

Next, each peak coinciding with the date of a milestone described by Wong and Dlugosz (2014) was further inspected to determine the most referenced articles from that year. As noted in our methods section, RPYS analysis was determined to successfully identify a milestone article if (1) the spectrogram revealed a peak on the corresponding milestone year and (2) the milestone articles noted by Wong and Dlugosz were amongst the 10 most referenced articles from that year. These results are detailed in Table 2. Nearly every peak year corresponding to a milestone year contained a milestone article as the top referenced article.

Milestones (Wong & Dlugosz, 2014; Figure 2)	Milestone Article (Wong & Dlugosz, 2014; bibliography)	RPYS Peak	Rank in Search			
			A	B	C	D
1827: First clinical description of BCC	Jacob (1827). <i>Dublin Hosp Rep</i>	Yes	1	1		
1950s: One eyed lambs in Idaho	Binns et al (1959) <i>J Am Vet Med Res</i> ^a	Yes			1	
1960: Nevoid BCC Syndrome described	Gorlin & Goltz (1960). <i>N Engl J Med</i>	Yes	1	1		
1963: Corn lily is teratogenic	Binns et al (1963). <i>Am J Vet Res</i>	Yes	1	1	1	1
1968: Teratogen in corn lily identified: cyclopamine	No Document Referenced ^b	Yes	1	1	1	1
1980: Hh & other embryonic patterning genes discovered	Nusslein-Volhard, Wieschaus (1980). <i>Nat</i>	Yes	1	1	1	1
1992: <i>Drosophila</i> hh gene cloned	No Document Referenced	No				
1993: Mammalian Hh genes cloned	No Document Referenced ^c	Yes				1
1996: <i>PTCH1</i> mutations in NBCCs, BCC	Hahn et al (1996). <i>Cell</i>	Yes	1	1	2	1
	Johnson et al (1996). <i>Science</i>	Yes	2	2	4	2
1996: Ptch is a Hh receptor	Stone et al (1996). <i>Nat</i>	Yes	4	4	6	5
1996: Cyclopia by impaired Hh signaling	Chiang et al (1996). <i>Nat</i>	Yes	6	5	1	6
1996: HH pathway mutations in human cyclopia	Belloni et al (1996). <i>Nat Genet</i>	Yes			7	9
	Roessler et al (1996). <i>Nat Genet</i>	Yes	10		3	7
1997: First transgenic BCC mouse model	Oro et al (1997). <i>Science</i>	No				
1997: <i>Ptch1</i> mutant mice phenocopy NBCCs	Goodrich et al (1997). <i>Science</i>	No				
1998: Cyclopamine inhibits Hh signaling	Cooper et al (1998). <i>Science</i>	Yes		2		2
1998: SMO mutations in BCC	Xie et al (1998). <i>Nature</i>	Yes		1		1
2000: Cyclopamine inhibits oncogenic Hh signaling	Taipale et al (2000). <i>Nature</i>	Yes			1	
2002: First synthetic small molecule Hh pathway inhibitor	Chen et al (2002). <i>Genes Dev</i>	Yes	1		1	1
2003: Multiple cancers with aberrant Hh signaling	Grachtchouk et al (2003). <i>EMBO J</i>	No				
2003: SMO antagonist pre-clinical BCC Tx	No Document Referenced	No				
2004: SMO antagonist pre-clinical medulloblastoma Tx	No Document Referenced ^d	Yes		2	7	
2009: Phase I trials of systemic Hh pathway inhibitor	Rudin et al (2009). <i>N Engl J Med</i>	Yes	2	2	2	2
	Von Hoff et al (2009). <i>N Engl J Med</i>	Yes	1	1	1	3
	Yauch et al (2009). <i>Science</i>	Yes	3	3	3	1
2011: Topical Hh pathway inhibitor BCC trials	Skvara et al (2011). <i>J Invest Dermatol</i>	No				
	Tang et al (2011). <i>Clin Cancer Res</i>	No				
2012: Hh inhibitor vismodegib FDA approved	No Document Referenced ^e	Yes		1		
2012: Systemic Hh inhibitor BCC prevention trial	Tang et al (2012). <i>N Engl J Med</i>	Yes		2		

- a Wong and Dlugosz do not include a reference to any articles published in the 1950s. They do, however, reference a 1978 review by Keefer, who in turn cites Binns et al (1959) to describe operators experiencing up to 25% deformed lambs in bands of ewes from areas of Idaho.
- b Though Wong and Dlugosz list 1968 as a milestone which is detected by RPYS, they do not reference the 1968 article on cyclopamine by Keeler and Binns. They do, however, cite Keeler and Binns (1966), which identified three alkaloids potentially inducing cyclopia.
- c Wong and Dlugosz include 1993 as a milestone which is detected by RPYS. They do not reference any 1993 article in their bibliography, however, RPYS analysis identifies an article cloning of a family of mammalian hedgehog genes by Echelard et al (1993) as the most referenced paper in 1993.
- d Wong and Dlugosz include 2004 as a milestone which is detected by RPYS. However, they do not reference any 2004 article in their paper. RPYS analysis of the second WoS search identifies an article by Romer et al (2004) on suppressing sonic hedgehog pathway to eliminate medullablastomas in mice.
- e FDA approval follows positive results from clinical trials. While there was no formal citation for FDA approval, Wong and Dlugosz do cite the 2012 publication of Clinical Trials on Vismodegib by Sekulic et al. (2012).

Table 2. Performance of RPYS in detecting seminal articles compared to key research breakthroughs described by subject matter experts. The leftmost column, Milestones, provides the year and description of key milestones noted in Wong and Dlugosz (2014; figure 2). The next column, Milestone Articles, maps articles from the bibliography or text of Wong and Dlugosz to the milestones they describe. We note that for some occasions multiple papers map to milestones. The column RPYS Peak reveals whether or not RPYS results of searches A, B, C or D yield a peak in the spectrogram corresponding to that milestone year. Finally, the last set of columns, Rank in Search, shows for each search A--D whether the milestone articles underlying an RPYS peak was in the top 1--10 most cited references for that year.

Further, RPYS aided in the clarification of seminal research underlying some of the milestones discussed in Wong and Dlugosz that were not explicitly cited (see Table 2 footnotes for more details). For example, Wong and Dlugosz note the 1950s occurrence of “one eyed lambs in Idaho” as a seminal event for research into the underlying mechanisms of basal cell carcinoma. However, the article does not cite any specific works published in the 1950s. It does, however, refer to a review article on this topic:

The identification of the first pharmacological inhibitor of Hh signaling was based on a series of pivotal observations and discoveries dating back to the 1950s (Figure 2), when up to 25% of lambs on certain Idaho ranches were born with severe craniofacial deformities, including a single eye (cyclopia) (reviewed in Keeler, 1978).

Examining the first paragraphs of the cited review by Keeler (1978) reveals the following description:

During the first half of the century, a significant percentage of lambs born in certain parts of Idaho were congenitally deformed (1-3)... By the 1950s some operators were experiencing an incidence of up to 25% deformed lambs in bands of ewes from these areas (1-3).

In this section, Keeler cites a 1959 article entitled “A congenital cyclopean--type malformation in lambs” by Wayne Binns et al. This same article is the most referenced article underlying a peak in 1959 in the spectrogram for Search C. In such cases, we consider the results of RPYS to have successfully captured a milestone article.

5. Conclusions

We demonstrate the application of new analytic techniques to connect major scientific discoveries with medical innovation. We show that multi--RPYS and standard RPYS, a relatively new class of citation algorithms, successfully capture key milestones in a field. We demonstrated that the combination of multi--RPYS and inferential statistics far exceeded chance expectations for identifying milestone years (70% success rate compared with less than 10% expected by chance). Then, we showed that 76% of the milestone papers underlying milestone years discussed by experts from the field of cutaneous malignancy are easily identifiable using online--available standard RPYS tools. This offers the possibility that recently developed RPYS tools (Comins and Leydesdorff 2016b; Leydesdorff et al. 2014; Thor et al. 2016) could empower a broader array of researchers to discover seminal works in scientific areas.

We wish to emphasize that expert opinion is necessary for contextualizing and describing the seminal works; but RPYS can expedite the process of finding milestones in the literature. In short, we conclude that RPYS represents a significant advance for maturing citation algorithms to identify seminal works in scientific fields, a key goal in scientometrics.

The performance achieved by these algorithms also opens the door for broader use of data--driven tools for mapping major milestones in the history of medical innovation (such as the prototype described in Comins & Leydesdorff, 2016b). Indeed, with the astounding pace of scientific growth (Bornmann & Mutz, 2015) and the tendency for fields with rapidly changing research fronts to cite more recent literature (Leydesdorff et al., 2014), such tools provide an easier way to identify key scientific activities influencing areas of medical innovation. RPYS and multi--RPYS offer a general-purpose method to revive the appreciation of early scientific achievements for the next generation of scientists and engineers, policy makers, funding agency program managers as well as lay people alike.

References

- Agarwal, P., & Searls, D. B. (2009). Can literature analysis identify innovation drivers in drug discovery? *Nature Reviews Drug Discovery*, 8(11), 865-878.
- Bornmann, L., & Mutz, R. (2016). Growth rates of modern science: A bibliometric analysis based on number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215--2222.
- Comins, J. A., & Hussey, T. W. (2015a). Compressing multiple scales of impact detection by Reference Publication Year Spectroscopy. *Journal of Informetrics*, 9(3), 449--454. doi:10.1016/j.joi.2015.03.003

- Comins, J. A., & Hussey, T. W. (2015b). Detecting seminal research contributions to the development and use of the global positioning system by reference publication year spectroscopy. *Scientometrics*, 104(2), 575–580. doi:10.1007/s11192-015-1598-2
- Comins, J. A., & Leydesdorff, L. (2016). RPYS i/o: software demonstration of a web-based tool for the historiography and visualization of citation classics, sleeping beauties and research fronts. *Scientometrics*, 107(3), 1509-1517.
- Comins, J. A., & Leydesdorff, L. (2016). RPYS i/o: software demonstration of a web-based tool for the historiography and visualization of citation classics, sleeping beauties and research fronts. *Scientometrics*, 107(3), 1509-1517.
- Comins, J. A., & Leydesdorff, L. (2016a). Identification of long-term concept-symbols among citations: Can documents be clustered in terms of common intellectual histories? *ArXiv*, 1601.00288, 1–20.
- Comins, J. A., & Leydesdorff, L. (2016b). RPYS i/o: A web-based tool for the historiography and visualization of citation classics, sleeping beauties, and research fronts. *Scientometrics*, 1–9. doi: 10.1007/s11192-016-1928-z
- de Solla Price, D. J. (1965). Networks of scientific papers. *Science*, 149(3683), 510–515. <http://www.eigenfactor.org/methods.php/methods.pdf>. Accessed 16 December 2014
- Elango, B., Bornmann, L., & Kannan, G. (2016). Detecting the historical roots of tribology research: A bibliometric analysis. *ArXiv*, 1601.00141.
- Garfield, E., Malin, M. V., & Small, H. (1978). Citation data as science indicators. In Y. Elkana, J. Lederberg, R. K. Merton, A. Thackray & H. Zuckerman (Eds.), *Toward a metric of science: the advent of science indicators* (pp. 179--207). New York, etc.: Wiley.
- Garfield, E., Sher, I. H., & Torpie, R. J. (1964). *The use of citation data in writing the history of science*. Philadelphia, PA: Institute for Scientific Information.
- Garfield, E., Sher, I. H., & Torpie, R. J. (1964). *The use of citation data in writing the history of science*. Philadelphia, PA: Institute for Scientific Information.
- Kostoff, R., & Shlesinger, M. (2005). CAB: citation-assisted background. *Scientometrics*, 62, 199–212. <http://www.akademai.com/index/R56K1LR0H3605886.pdf>. Accessed 16 December 2014
- Leydesdorff, L., & Rafols, I. (2011). How Do Emerging Technologies Conquer the World? An Exploration of Patterns of Diffusion and Network Formation. *Journal of the American Society for Information Science and Technology*, 62(5), 846-860.
- Leydesdorff, L., Bornmann, L., Comins, J. A., Marx, W., & Thor, A. (2016). Referenced Publication Year Spectroscopy (RPYS) and Algorithmic Historiography: The Bibliometric Reconstruction of Andrés Schubert's Œuvre. Pp. 95--112 in: W. Glänzel (Ed.), *Festschrift for Andras Schubert*, Louvain: International Society for Scientometrics and Informetrics ISSI.
- Leydesdorff, L., Bornmann, L., Marx, W., & Milojević, S. (2014). Referenced Publication Years Spectroscopy applied to iMetrics: *Scientometrics*, *Journal of Informetrics*, and a relevant subset of *JASIST*. *Journal of Informetrics*, 1901, 1–34. <http://www.sciencedirect.com/science/article/pii/S1751157713001077>. Accessed 23 September 2014
- Marx, W., & Bornmann, L. (2013). Tracing the origin of a scientific legend by reference publication year spectroscopy (RPYS): the legend of the Darwin finches. *Scientometrics*, 99(3), 839–844. doi:10.1007/s11192-013-1200-8
- Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751–764. doi:10.1002/asi.23089
- Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751-764.
- Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751-764.
- Mina, A., Ramlogan, R., Tampubolon, G., & Metcalfe, J. S. (2007). Mapping evolutionary trajectories: Applications to the growth and transformation of medical knowledge. *Research Policy*, 36(5), 789-806.
- Mogoutov, A., Cambrosio, A., Keating, P., & Mustar, P. (2008). Biomedical innovation at the laboratory, clinical and commercial interface: A new method for mapping research projects, publications and patents in the field of microarrays. *Journal of Informetrics*, 2(4), 341-353.
- Nelson, R. R., Buterbaugh, K., Perl, M., & Gelijns, A. (2011). How medical know-how progresses. *Research Policy*, 40(10), 1339--1344.
- Petersen, A., Rotolo, D., & Leydesdorff, L. (2016). A Triple Helix Model of Medical Innovations: Supply, Demand, and Technological Capabilities in Terms of Medical Subject Headings. *Research Policy*, 45(3), 666-681. doi: 10.1016/j.respol.2015.12.004
- Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the ...*, 105(45), 17268–17272. <http://www.pnas.org/content/105/45/17268.short>. Accessed 16 December 2014
- Swanson, D. R. (1990). Medical Literature as a Potential Source of New Knowledge. *Bull. Med. Libr. Assoc.* 78, 29--37.
- Thor, A., Marx, W., Leydesdorff, L., & Bornmann, L. (2016). Introducing CitedReferencesExplorer (CRExplorer): A program for Reference Publication Year Spectroscopy with Cited References Disambiguation. *Journal of Informetrics*, 10(2), 503-- 515. doi: 10.1016/j.joi.2016.02.005
- Van Raan, A. (2000). On growth, ageing, and fractal differentiation of science. *Scientometrics*, 47, 347–362. <http://link.springer.com/article/10.1007/s11192-014-1465-6>. Accessed 16 December 2014
- Von Hippel, E. (1988). *The Sources of Innovation*. New York/Oxford: Oxford University Press.
- Williams, R.S., Lotia, S., Holloway, A.K., & Pico, A.R. (2015). From scientific discovery to cures: Bright stars within a galaxy. *Cell*, 163, 21--23.
- Wong, S. Y., & Dlugosz, A. a. (2014). Basal cell carcinoma, Hedgehog signaling, and targeted therapeutics: the long and winding road. *The Journal of Investigative Dermatology*, 134(e1), E18–22. doi:10.1038/skinbio.2014.4
- Wray, K. B., & Bornmann, L. (2015). Philosophy of science viewed through the lense of “Referenced Publication Years Spectroscopy”(RPYS). *Scientometrics*, 102(3), 1987--1996.