Can Numbers of Publications on a Specific Topic Observe the Research Trend of This Topic: A Case Study of the Biomarker HER-2?

Yuxian Liu[1,2,3], Michael Hopkins[2] and Yishan Wu[4]

yxliu@tongji.edu.cn, m.m.hopkins@sussex.ac.uk, wuyishan@istic.ac.cn

[1]Tongji University, Tongji University Library, Siping Street 1239, 200092 Shanghai (China);

[2]Tongji University Development & Planning Research Center, Siping Street 1239, 200092 (China);

[3]Univ. of Sussex, School of Business, Management and Economics, SPRU, Falmer BN1 9SL, Brighton (UK)

[4]Institute of Scientific and Technical Information of China, 15 Fuxinglu, Beijing 100038 (China)

Abstract

Using the accumulative publication data on HER-2 and its trend line, we draw the accumulative curve of the publication data. We discuss the characteristics of the accumulative publication curve, and how these characteristics change with respect to the different trend lines. We find that the points that regression line and the publication curve intersect with each other and the minimum points with respect to the trend lines do not change very much in both exponential trend line and linear trend line even if the exponential trend line raises itself much faster than the linear trend line. These data points are formed around the time when the significant discoveries are made and the related regulations are executed. These significant discoveries and regulations impact how and where the research should go and how the basic discoveries influence their application. The accumulative publication curve itself tells us very little about science. However the change of the accumulative publication curve with respect to the trend lines may tell us how science evolves. The content in the publications with significant scientific value may change the direction and trend of research, while research may change the publication trend the other way round. We may say that important scientific discoveries and regulations on clinical practice act as tipping points or act as drivers of change in the rates of scientific publications on the topic of HER-2. This induces us further to explore how scientific events drive the publication process. We may expect that through the publication process, we can monitor the scientific process.

Conference Topic

Theory

Introduction

The number of publications is widely used to measure the output or the productivity of researchers or their affiliated institutes. Hence, it is also used to compare the output of different countries (Bornmann & Marx, 2013; Zhu et al., 2004; Inglesi-Lotz & Pouris, 2011; Garfield, Pudovkin, & Paris, 2010). (China is ranked the second in terms of output of scientific research measured by the number of publications.) It is normally regarded as a quantitative indicator. The number of citations is supposed to measure the impact or the visibility of the researchers or their affiliated institutes that are investigated (Garfield, 1955). Sometimes it is even referred to as the indicator that measures the quality of the research in the cited article that a researcher has performed.

However, these measurements arouse a heated debate. In the December 16, 2012, the concerned scientists gathered in the Annual Meeting of the American Society for Cell Biology developed a set of recommendations referred to as the *San Francisco Declaration on Research Assessment* (DORA). DORA aimed to stop the use of the "journal impact factor" (JIF) in judging an individual scientist's work. They invited interested parties to indicate their support by adding their names to this declaration. Later the editor-in-chief of *Science* Bruce Alberts published an editorial to support this declaration. He thought the evaluation based on JIF was destructive and just encouraged "me-too science" and hence blocked innovation and created a strong disincentive to pursue risky and potentially groundbreaking work. Many leading scientists and scientific organization endorsed in this declaration (Alberts, 2013). JIF, a scientometric indicator based on the number of publications and the number of citations,

was originally created as a tool to help librarians to select journal to purchase, but later it is frequently used as a measure of the scientific quality of research in an article published in this journal and act as the primary parameter with which to compare the scientific output of individuals and institutions. Some academic institutes even use it to decide if a researcher should be funded or promoted as a tenure member (Garfield, 1999; Alberts, 2013). However, this practice arouses the fierce objection by scientists who are evaluated.

Bibliometricians also gave their voices to this phenomenon. Wouters, Glänzel, Gläser, & Rafols (2013) call for the urgent debate on the dilemmas of performance indicators of individual researchers. The Higher Education Funding Council for England (HEFCE), which distributes public money for higher education to universities and colleges in England and ensures that this money is used to deliver the greatest benefit to students and the wider public, carry out a work to review the role of metrics in the assessment and management of research. (http://www.hefce.ac.uk/whatwedo/rsrch/howfundr/metrics/). In the review, the working group launched a call for evidence to gather views and evidence relating to the use of metrics in research assessment and management. Elsevier and SPRU responded to the call. Ismael Rafols, Paul Wouters and Sarah de Rijcke organized a special session on the quality standards for evaluation indicators: Any chance for the dream to come true? (STI program). This session initiated to make the Leiden manifesto on the research assessment. van Raan, a scientometrics pioneer and gatekeeper (Garfield, Pudovkin, & Paris, 2010), will coordinate among different aspects so that this manifesto could be accepted widely. All these principles and responses, without exception, mention that quantitative information provided by metrics must be complemented by qualitative evidence to ensure the most complete and accurate input to answer a question. Even DORA recommended that the funding agencies should consider a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice. DORA also recommended the publishers should make available a range of article-level metrics to encourage a shift toward assessment based on the scientific content of an article (DORA).

Garfield (1979, p. 62) illustrated that:

> "If the literature of science reflects the activities of science, a comprehensive, multidisciplinary citation index can provide an interesting view of these activities. This view can shed some useful light on both the structure of science and the process of scientific development."

However, can metrics drawn from publications and citations provide qualitative indicators that reveal the contents of the publications so that metrics can measure the way the contents of the publications influence policy and practice? Liu & Rousseau (2013, 2014) expounded that citation in essence is the interaction of the perspectives on a specific scientific phenomenon, hence can be used to reveal how the scientific phenomenon is understood. With the help of the regression line and a detrended curve, Liu & Rousseau (2012) show that the citation diffusion curve of an article containing a really original idea has an S-shape similar to the standard innovation diffusion curve. The convex part corresponds to the academic phase of the field that Kao's idea initiated, while the concave part corresponds to the technology dominated phase. The curve in the post-technology phase paralleled the regression line. The points of inflection correspond to the phase transition from academic to application research, while minima indicate a breakthrough in academic phase, and maxima indicate a breakthrough in the technology dominated phase. This implies that breakthroughs may directly influence the rate of change of the diffusion process while phase transfers may influence the rate of change implicitly. They claimed that the theory of diffusion process expounded in this article have the potential use of discerning breakthrough and turning points in an S & T area and finding social, technological, political and economic factors influencing the development of science. Can we use the number of publications on a specific topic to

observe the research trend? How the regression lines and the detrended forms of the publication curve tell us about the development of science? Can we discern the breakthrough and turning points between the academic phase and applied phase? Can we find social, technological, political and economic factors influencing the development of science? In this article, we will use the publications on Biomarker Her 2 to illustrate how the scientific activities on a specific research topic influence the publication process. With the help of the regression line and the detrended forms of the publication curve, we try to identify the breakthrough in this area and trajectory of translating research finding into diagnostic tools, medicines, procedures, policies and education. We will combine descriptive material on the development of the research domain with the publication growth - presents a model of interconnections of the publication and citation process, we analyze the cumulative publication curve and compare it to major events in the field. We will show that important scientific discoveries and regulation of clinical practice act as tipping points/ drivers of change in the rates of scientific publications on the topic of HER-2.

Data

After comprehensive literature research, we determined our search string:
TS=("CerbB2*" OR "CerbB-2*" OR "Cer-bB2*" OR "C-erbB2*" OR "Cer-bB-2*" OR "C-erbB-2*" OR "C-er-bB2*" OR "C-er-bB-2*" OR "Cerb B2*" OR "Cerb B 2*" OR "erbB2*" OR "erbB-2*" OR "er-bB2*" OR "er-bB-2*" OR "erb b2" OR "erb b 2" OR "HER2" OR "Epidermal growth factor receptor 2" OR "EGFR2" OR "CD340" OR "her 2")
These words include all the spelling variants related to the biomarker Human Epidermal Growth Factor 2. Among these words, “Her 2” is the only word that is not specific which may bring us some noising results because “her 2” can be used as in “her 2 children” which has nothing to do with Human Epidermal Growth Factor 2. Worse, children can be replaced by any nouns. Between her 2 and the nouns, any adjectives can be added in between. Even worse, since the Web of Science (WoS) ignores all punctuations, any punctuations can be added in between. Also one item that has “her 2 children” does not necessarily mean it is not what we need. Even the articles which deal with Human Epidermal Growth Factor 2 do not exclude the expression “her 2 children”. These situations make it very difficult for us to formulate an effective search string. However, we use the position information and its follow up to judge if these articles are related to the topic that we are searching by a program (Chavarro & Liu 2014, Lang, Liu & Chavarro, 2015), if it cannot be judged by a program, we judge it manually. We have got 98 articles that are not related to our topic. We downloaded all these data in 27 May 2014 and then excluded these 98 articles. Hence we get 30,056 articles. Since the gene of Her2/neu did not have a uniform name at the beginning when the scientists found this gene, we picked up some articles from the reference list of the early articles. And we exclude the articles published in 2014, and then we get 29,210 publications. Using these 29,210 records we do some bibliometric analysis.
The numbers of publications per year increase in roughly linearly. It is said that when a research topic turns to the application science, fewer and fewer publications will be published, instead, more and more patents will be approved. But in our case, it is the opposite, the research topic on HER-2 has already been applied in the diagnosis and therapy, the numbers of the publications on this topic do not decrease at all.

Table 1. Cumulative numbers of publications, the first and the second order differences.

Year	*cumulative numbers of publication*	*the first order difference*	*the second order difference*
1981	1	1	
1982	1	0	-1
1983	1	0	0
1984	3	2	2
1985	6	3	1
1986	12	6	3
1987	29	17	11
1988	68	39	22
1989	133	65	26
1990	261	128	63
1991	467	206	78
1992	763	296	90
1993	1126	363	67
1994	1581	455	92
1995	2046	465	10
1996	2530	484	19
1997	3048	518	34
1998	3624	576	58
1999	4312	688	112
2000	4996	684	-4
2001	5980	984	300
2002	7006	1026	42
2003	8141	1135	109
2004	9414	1273	138
2005	10922	1508	235
2006	12527	1605	97
2007	14196	1669	64
2008	16262	2066	397
2009	18633	2371	305
2010	21040	2407	36
2011	23500	2460	53
2012	26423	2923	463
2013	29210	2787	-136

Methodology: Regression Trend Lines and Detrended Curves of Time Series Data

Table 1 is a time series data. A time series is a sequence of data points, typically consisting of successive measurements made over a time interval. In informetrics, the time interval can be defined in different shift (Liu & Rousseau, 2008). Normally we make a scatter diagram to see whether data change linearly or nonlinearly. Then we make a regression analysis to find the best-fitting curve to see how the data change over time. We can get a regression equation to explain the

degree of association or the relationship between the data and time. Based on the equation that fits past data as well as possible, we can predict values of the variable at points other than the observation points.

The linear regression is the straight line. The curves of the nonlinear regression curves, depended on the regression equations, have different shapes. For example, the curve can be exponent curves if the regression equation is exponent function. The other possible curves can be logarithmic curve, power curve and multinomial curve. The straight line from the linear regression and the curve from the nonlinear regression are also called trend lines. Figure 1 show the exponential, multinomial, power, linear and logarithmic regression curves of data in Table 1.

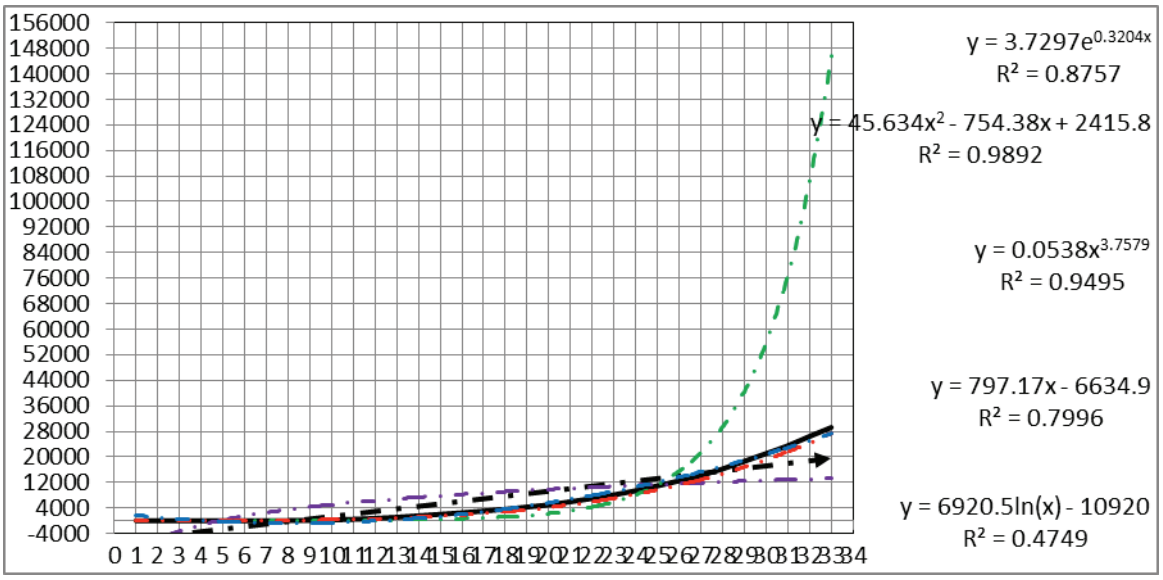

Figure 1.The exponential, multinomial, power, linear and logarithmic regression curves of data in Table 1.

Detrended curve (Shiavi, 1991) is a detrended description of the data. In order to draw a detrended curve, we find a trend line first and then calculating the difference between the overserved data and the trend line. It will give us a view on how the data change in terms of trend line. Peng et al. (1994, 1995) introduce the detrended fluctuation analysis. It is a scaling analysis method used to estimate long-range temporal correlations form. In other words, if a sequence of events has a non-random temporal structure with slowly decaying auto-correlations. It hence can eliminate the trend that self-affinity. By discerning long range correlation, it can help us understand what dominates the change of the data in the time series.

In this article, instead of calculating the difference between the observed data and the trend line, we will rotate abscissa to the paralleling line of the regression line and make the line touching the edge of the scatter diagram. The ordinate will pass through the first observation point so that all the numbers are positive. We then establish a new coordinate system. We will see how the data change with respect to the regression line

Results

We can choose different regression trend lines. In this article, we choose the best fitted straight line. Figure 2 shows the cumulative curve of the numbers of publications on her 2, its regression line and its minimum with respect to the regression line. We can see that the cumulative curve of the numbers of the publications on HER-2 is convex. The regression line intersects with the original data around 1987-1988 and 2007-2008. The minimum with respect to the regression line is around 1998-1999 (1 is the year 1981, 2 is 1982 and so on).

Now we know gene HER-2 was identified in 1981 by transfection studies with DNA from chemically induced rat neurogliobalstomas by Shih, Padhy, Murray and Weinberg (1981). From 1981 to 1987, several groups identified this gene independently (Schechter et al., 1985; Coussens et al., 1985; Semba, Kamata, Toyoshima, & Yamamoto, 1985; Fukushige et al., 1985). Slamon, Clark, Wong, Levin, Ullrich, and McGuire (1987) found correlation of relapse and survival with amplification of the HER-2 oncogene. HER-2 became a significant prognostic factor. Since then Slamon started to do research on binding to the HER-2 protein and prevents it from relaying a signal that stimulates the cancer cell to divide (Pioneers, 2007). In 1998, Herceptin was approved by FDA. Since then a revolutionary treatment started its journey in the history of human being to conquer the disease, based on the gene analysis, personalized treatment appear in the horizon that people can see. In 2007, American Society of Clinical Oncology (ASCO) and The College of American Pathologists (CAP) developed guidelines for when and how the status of HER-2 should be tested (Wolff et al., 2007). This guidelines were updated in 2013 (Wolff et al., 2013). Since then the test for the statues of HER-2 and clinical treatment with Herceptin become a standard test and treatment. However, as Herceptin did not take effect in some patients, the subpopulation remains to be defined, and side effects including cardiotoxicity need to be solved (Kumler, Tuxen, & Neilsen, 2014), HER-2 is still a topic that needs more investigations. We indicate these important events in Figure 2.

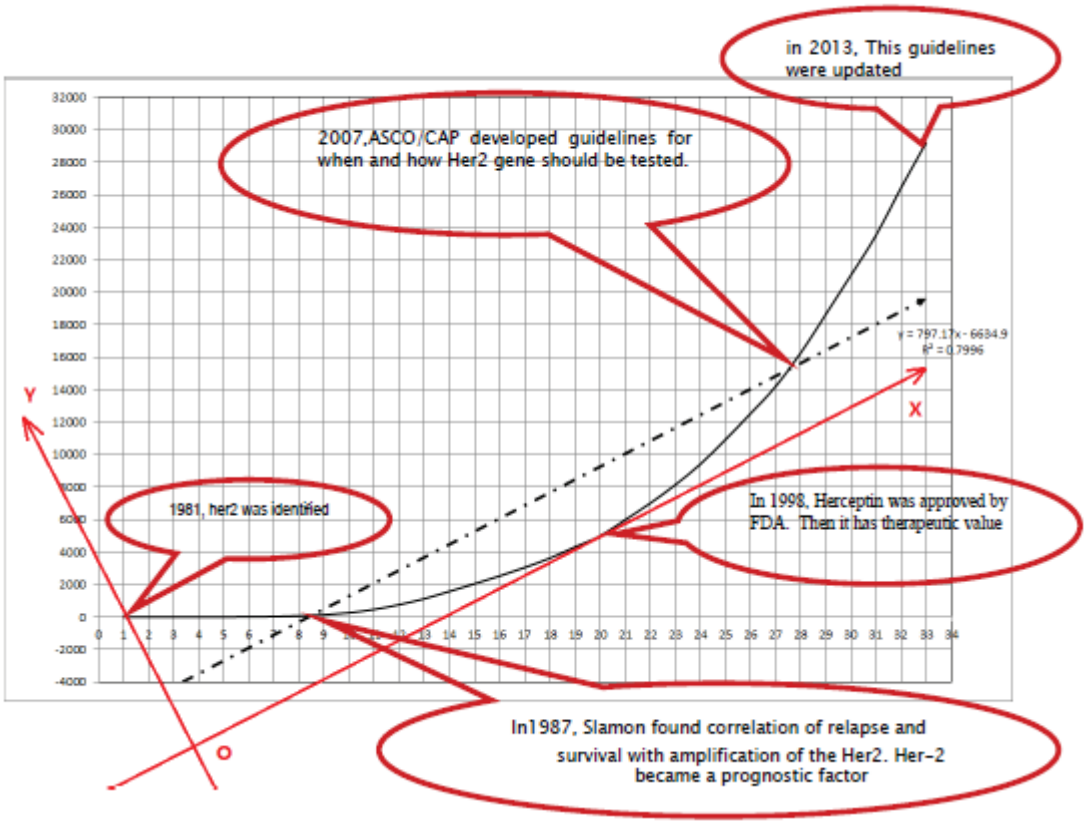

Figure 2. The cumulative curve of the numbers of publications on her 2, its regression line and its minimum with respect to the regression line.

We know that the minimum is a key point where the first-order derivative changes from negative to positive. If a curve shows the status of a thing that changes over time, we say it describes a kind of motion. The motion described by this curve changes from decreasing to increasing in the minimum point. A motion may have a different appearance as viewed from a different reference frame. If we choose the actual data as reference frame, to see how the trend changes, we can see that the discovery of the correlation of relapse and survival with amplification of her-2 oncogene in 1987 changes the trend reflected with publication data. This discovery made the amplification of her-2 a significant predictor and prognostic factor.

The numbers of publications start to increase significantly, the passion on this research topic is activated. Though with respect to the trend line, the original data curve decreases monotonically; the curve did not begin to increase until 1998 when Herceptin was approved by FDA. It is similar to the minimum point in cumulative number of citations curve of Kao when optical fiber was invented by Corning Glass Works in 1970. The crucial material problem, optical fiber, which Kao said in his conclusion "appears to be one, which is difficult but not impossible" was solved. This invention helped Kao realized his dream that no one believed it at the beginning (Liu & Rousseau, 2012). It is a coincidence that no one believed that the method Dr. Slamon used would work and the drug he created would be approved by FDA. On the contrary, everyone thought Dr. Slamon was crazy and he could not even find a student assistant majoring in science at the beginning (see the movie: Living proof and Bazell, 1998). In 2007, HER-2 test in breast cancer was recommended by ASCO-CAP, HER-2 research entered into another stage. The second order difference decreases after 2008. It dropped tremendously in 2010 and 2011. But in 2012, it went up tremendously which probably was caused by the fact that the recommendation guideline was challenged by the clinic practices and the new progresses. In 2013, ASCO/CAP convened an Update Committee that included coauthors of the 2007 guideline to conduct a systematic literature review and update recommendations for optimal HER-2 testing. In 2013, the second order difference become negative. Does the curve reach the point of inflection? We know the negative second order difference means the curve change from convex to concave. So far we cannot get to this conclusion. More observations are needed, at least we need to know how many publications on HER-2 will be published in 2014 so that we can judge whether it is an innate trend or just an occasional fluctuation. However, since major debate was settled down, though HER family oncogene (erbb1 erbb2, erbb3, erbb4) need to be dually blocked, and relative subpopulation needed to be defined and side effects refrain the use of some new developed medicine. For the moment there is an urgent need for prospective biomarker-driven trials to identify patients for whom dual targeting is cost effective (Kumler, Tuxen, & Neilsen 2014), we say it is not a major obstacle. We expect that the year when the breakthrough will make on these obstacles will appear in the maximum point on the curve drawn by the numbers of the publications on the HER-2. But it would depend on whether the research topic HER-2 gives rise to the other research topic.

The predictive, prognostic and therapeutic value of HER-2 are what changes the trend of research. The discoveries of these values of HER-2 influence the diffusion of the knowledge on HER-2 in the landscape of human intellectual space.

Selection of Trend Lines and the Different Implications that Detrended Line can Give Us

We can choose exponential, linear, logarithmic or power function as the trend lines to see what the data can tell us. Intuitively these trend lines are totally different, we hence imagine that the different trend lines can tell us totally different stories. But Figure 1 tells us the points that the different trend lines cross the data are slightly different, all around 2005-2008 even if the exponential trend is a much faster trend than the linear one. However, it is difficult to establish a new coordinate system to see clearly what the data tells us. Since the exponential curve is a straight line in semi-logarithmic system, we draw a scatter diagram in semi-logarithmic system (Figure 3). The data curve is concave upwards with respect to the exponential trend line. We can see the extremum with respect to the exponential trend line is around 1994, a little bit earlier than the time when the Herceptin was approved. However, it is in 1994 that Prof. Slamon finished phase 3 trial and was waiting for the decision of FDA. The first point that the trend line crossed the data is the same, but the second point is a little bit earlier. But the 2007 guideline was accepted for publication in September 27, 2006. The

expert panel was convened in 2005 and started to work on the guideline. It seems as if the shift of time is still in the acceptable region.

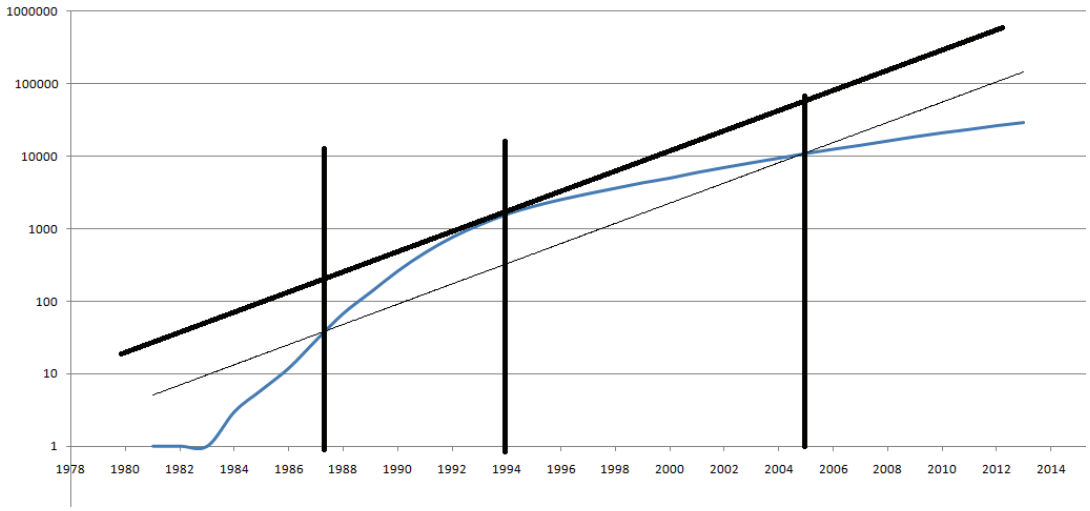

Figure 3. Cumulative publication data curve and its exponential trend line in the semi-logarithmic framework.

Publication and Citation Diffusion Process

Liu (2011) and Liu and Rousseau (2012, 2013, 2014) explore the determining factors that influence the citation process, and link the citation to the cognitive process of a scientific phenomenon under investigation. Through these articles, we illustrated the interaction of different perspectives on the phenomenon under investigation and how it is that the new ideas are accepted by academia determine the citation diffusion process.

In this investigation, we show that publication data curve with respect to the trend line can reflect how the important scientific events such as scientific discoveries and the release of government regulations in the clinical practice can change the trends of the publication process. Obviously, the primary knowledge creation process influences not only the citation process but also publication process. The change of research trends can show themselves in the publication data curve with respect to the trend line.

Liu and Rousseau (2010) studied two forms of diffusion, namely diffusion by publication and by citation. They tried to illustrate that publication diffusion is dominated by the internal diffusion mechanism that originates from the fact that a group of scientists expands their own (field) border. The citation diffusion is dominated by the external diffusion mechanism that the publication of the group of scientists, published in more and more fields, have potential to be applied in the other fields. Obviously, the publication diffusion process and citation diffusion process are interlinked with each other in that publication diffusion process determines the citation diffusion process.

As a matter of fact, publication process is entangled with citation process. Figure 4 shows how these two processes are entangled. Once the scientist(s) are interested in the scientific phenomenon, on the one hand they observe this phenomenon and get some preliminary impressions, and from these impressions they formulate some scientific ideas. On the other hand, they read the literature, which discusses this phenomenon and the perspectives to interact with the ideas that they formed by their observations to help them to get new insight into the phenomenon, and they then begin to make a thorough investigation. From these investigations scientists get new perspectives on the phenomenon. They articulate the new perspectives into a publication. When they write the manuscript they cite the old perspectives in the literature (perhaps they also read the other literatures for new evidence to convince the

readers). Publication and citation are thus born. In this process, scientific phenomenon is more and more clearly cognized.

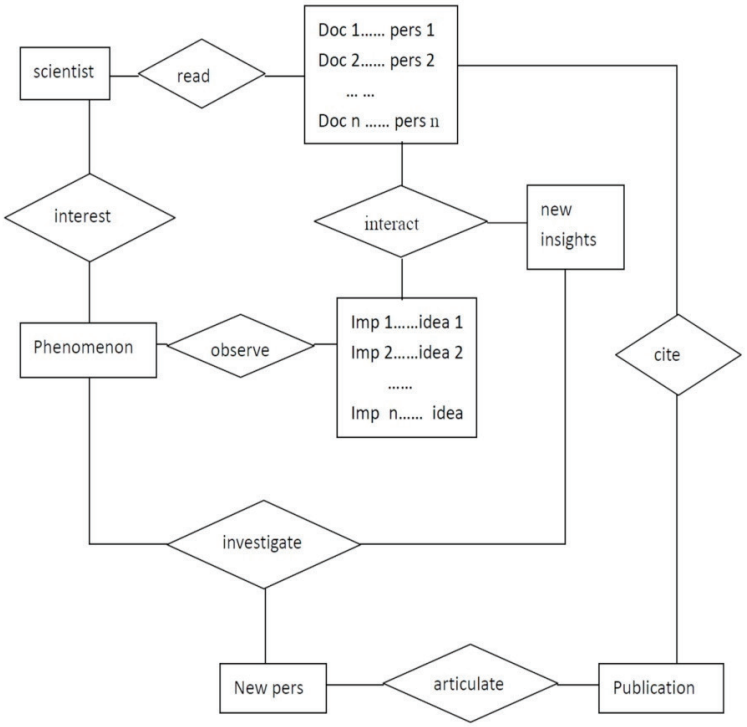

Figure 4. Entangled publication and citation diffusion process.

We can see from Figure 4 that the citation and publication processes are dynamic movement processes driven by the cognitive process of a phenomenon under investigation. The cognitive process is constituted (was led) by a series of scientific events. In this sense we may say that scientific events act as an engine to drive the evolution of science. Some events can lead the cognitive process to another direction. For example, the research in the publication Slamon, Clark, Wong, Levin, Ullrich and McGuire (1987) led HER-2 research from basic research to applied research. This event will change the research trend. Some events have no significant influence on the research trend.

Every scientific event could be represented by some publications on this event. In this sense, the scientific events drive the publication process, this process then drives the citation process. Scientific ideas that the publications convey are then diffused into the human intellectual landscape. So publication diffusion process may give us a deeper insight into the scientific events. The relationships between different publications are not as clear as that in citations, though through co-authorship or co-keywords we can establish different networks. But co-authorship or co-keywords did not reveal how the idea in one publication is diffused into the other publication. We cannot trace how the ideas in different publications interact with each other. Probably the mechanism of publication diffusion process needs to be explained via the citation diffusion process communicating different perspectives of the phenomenon under investigation. Therefore, citation and publication have a potential to reveal the cognitive process of the phenomenon under investigation.

However, we must understand how a scientific idea is diffused in the abstract intellectual landscape. This is the academic movement. In order to describe the academic movement we need to know where an idea comes from, where it will go, how fast the diffusion process is, how long is the distance from its start point to its destination. However, we face a lot challenges. First of all, we must mark the landscape with these scientific events. We have the

classification system such as the Library of Congress Classification System, Chinese Classification System, the WOS subject areas and the ESI fields. However, these systems alone cannot mark the scientific event. Because of the inaccuracy of this system, this kind of research does not give us more sense about the cognitive process of a research topic. Trochim and his colleagues (2011) proposed to identify "markers" in the translation process. They then assess the time that it takes for outputs to move across markers (Molas-Gallart, Este, Llopis & Rafols, 2014). Maybe this kind of mark system that embedded in a concrete scientific investigation will give us more information about the cognitive process of a scientific research.

Secondly, the distance in the human intellectual landscape may change over time and the destinations for the diffusion process are uncertain. These will make it very difficult to describe the scientific cognitive process via publication and citation diffusion process. These research questions deserve our effort. We would understand the scientific process more accurately if we could describe publication and citation diffusion processes more precisely. We can even anticipate what drives the evolution of science.

Conclusion

With the numbers of the publications on HER-2, we drew the accumulative curve of the publication data. We discuss the characteristics of the accumulative publication curve with respect to its trend lines and how its characteristics change in different trends. We find out the intersect points through regression line and the publication curve. These points are around the time when significant discoveries and regulations are made. These significant discoveries and regulations dominate how and where the research should go and how the basic discoveries influence their application. The accumulative publication curve itself tells us very little about how the science is evolving, but the change of the accumulative publication curve with respect to the trend lines may tell us more about the science. The content in the publication that has significant scientific value may change the direction and trend of research, hence change the publication trend reversely. We may say that important scientific discoveries and government regulations on clinical practice act as tipping points or act as drivers of change in the rates of scientific publications on the topic of HER-2. This makes us go further to explore how scientific events drive the publication process.

Acknowledgements

Yuxian Liu thanks Ismael Rafols, Raf Guns, Tim Engels, and Ronald Rousseau for the discussion in the early stage of this research. This work is supported by NSFC via 71173154. Yuxian Liu further acknowledges support from the China Scholarship Council.

References

Alberts, B. (2013). Impact Factor Distortions. *Science*, 340(6134), 787.

Bornmann, L. & Marx, W. (2013). Proposals of standards for the application of scientometrics in the evaluation of individual researchers working in the natural sciences. *Zeitschrift für Evaluation,* 12(1), 103-127.

Chavarro, D., & Liu, Y. (2014). How can a word be disambiguated in a set of documents: using recursive Lesk to select relevant records. http://www.gtmconference.org/pages/program.html

Coussens, L., Yang-Feng, T. L., Liao, Y. C., Chen, E., Gray, A., McGrath, J. et al. (1985). Tyrosine kinase receptor with extensive homology to EGF receptor shares chromosomal location with neu oncogene. *Science*, 230, 1132-1139.

Elsevier. (2014).Response to HEFCE's call for evidence: independent review of the role of metrics in research assessment. http://www.elsevier.com/__data/assets/pdf_file/0015/210813/Elsevier-response-HEFCE-review-role-of-metrics.pdf.

Fukushige, S. I., Matsubara K.I, Yoshida, M., Sasaki, M., Suzuki, T., Semba, K. et al. (1986). Localization of a novel v-erbB-related gene, c-erbB-2, on human chromosome 17 and its amplification in a gastric cancer cell line. *Molecular and Cellular Biology,* 6, 955-958.

Garfield, E. (1999). Journal impact factor: a brief review. *Canadian Medical Association Journal*, 161(8), 979-980.

Garfield, E (1979). *Citation Indexing.* New York: Wiley.

Garfield, E. (1955). Citation index for science. *Science*, 122, 108-111.

Garfield, E., Pudovkin, A. I., & Paris, S. W. (2010). A bibliometric and historiographic analysis of the work of Tony van Raan: a tribute to a scientometrics pioneer and gatekeeper. *Research Evaluation*, 19(3), 161-172.

Inglesi-Lotz, R., & Pouris, A. (2011). Scientometric impact assessment of a research policy instrument: the case of rating researchers on scientific outputs in South Africa. *Scientometrics*, 88(3), 747-760.

Kumler, I., Tuxen, M.K., & Neilsen, D. L. (2014). Anti-Tumour Treatment, A systematic review of dual targeting in Her-2 positive breast cancer. *Cancer Treatment Reviews*, 40, 259-270.

Lang, F., Liu, Y., & Chavarro, D. (2015). Improving accuracy in data collection: Can machine learning classification help? (in press)

Liu, Y.X. (2011). The diffusion of scientific ideas in time and indicators for the description of this process. Unpublished Doctoral Thesis, University of Antwerp, Belgium.

Liu, Y. X. & Rousseau, R. (2008). Definitions of time series in citation analysis with special attention to the h-index. *Journal of Informetrics,* 2(3), 202-210.

Liu, Y.X., & Rousseau, R. (2010). Knowledge diffusion through publications and citations: A case study using ESI-fields as unit of diffusion. *Journal of the American Society for Information Science and Technology*, 61(2), 340-351.

Liu, Y. & Rousseau, R. (2012). Towards a representation of diffusion and interaction of scientific ideas: the case of fiber optics communication. *Information Processing and Management,* 48(4), 791-801.

Liu, Y. & Rousseau, R. (2013). Interestingness and the essence of citation. *Journal of Documentation*. 69(4), 580-589.

Liu, Y. & Rousseau, R. (2014). Citation analysis and the development of science: a case study using articles by some Nobel Prize winners. *Journal of the American Society for Information Science and Technology,* 65(2), 281–289.

Peng, C-K., Buldyrev, S.V., Havlin, S., Simons, M., Stanley, H.E., & Goldberger, A.L. (1994). Mosaic organization of DNA nucleotides. *Physical Review E,* 49:1685-1689.

Peng, C-K., Havlin, S., Stanley, H.E., & Goldberger, A.L. (1995). Quantification of scaling exponents and crossover phenomena in nonstationary heartbeat time series. Chaos, 5:82-87.

Dr. Dennis Slamon's development of Herceptin revolutionized breast cancer treatment and accelerated research into therapies customized for each individual patient: the personal approach. (2007, Fall-Winter). *Triumph*, 13-16.

Schechter, A. L., Stern, D. F., Vaidyanathan, L., Decker, S. J., Drebin, J. A., Greene, M. I. et al. (1984). The neu oncogene: an erb-B-related gene encoding a 185,000-Mr tumour antigene. *Nature*, 312(5994), 513-516.

Semba, K., Kamata, N., Toyoshima, K., & Yamamoto, T. (1985). A v-erbB-related protooncogene, c-erbB-2, is distinct from the c-erbB-1/epidermal growth factor-receptor gene and is amplified in a human salivary gland adenocarcinoma. *PNAS,* 82 (19), 6497-6501.

Shiavi, R. (1991). *Introduction to Applied Statistical Signal Analysis*. Homewood, IL: Irwin, Aksen.

Shih, C., Padhy, L., Murray, M., & Weinberg, R. A. (1981). Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts. *Nature*, 290, 261-264.

Slamon, D. J., Clark, G. M., Wong, S. G. Levin, Ullrich, A., & McGuire, W. L. (1987). Human breast cancer: correlation of relapse and survival with amplification of the HER-2/neu oncogene. *Science*, 235, 177-182.

Truex, D., Cuellar, M., & Takeda, H. (2009). Assessing scholarly influence: using the Hirsch indices to reframe the discourse. *Journal of the Association for Information Systems,* 10 (7), 560-594.

Wolff, A.C., Hammond, M. E. H., Schwartz, J.N.,Hagerty, K.L., Allred, D.C., Cote, R. J. et al. (2007). American Society of Clinical Oncology/College ofAmerican Pathologists Guideline Recommendations for Human Epidermal Growth Factor Receptor 2Testing in Breast Cancer. *Arch Pathol Lab Med*, 131, 18-43.

Wolff, A.C., Hammond, M. E. H., Hicks, D. G., Dowsett, M., McShane, L.M. et al. (2013, November 1). Recommendations for Human Epidermal Growth Factor Receptor 2 Testing in Breast Cancer: American Society of Clinical Oncology/College of American Pathologists Clinical practice Guideline Update. *Journal of Clinical Oncology*, 31(31), 399-4013.

Wouters, P., Glänzel, W., Gläser, J. & Rafols, I. (2013). The dilemmas of performance indicators of individual researchers: an urgent debate in bibliometrics. *ISSI Newsletter*, 9(3), 48-53.

Zhu, X., Wu, Q., Zheng, Y. Z., & Ma, X. (2004). Highly cited research papers and the evaluation of a research university: A case study: Peking University 1974-2003. *Scientometrics*, 60(2). 237-247.