STI 2018 Conference Proceedings

Proceedings of the 23rd International Conference on Science and Technology Indicators

All papers published in this conference proceedings have been peer reviewed through a peer review process administered by the proceedings Editors. Reviews were conducted by expert referees to the professional and scientific standards expected of a conference proceedings.

Chair of the Conference

Paul Wouters

Scientific Editors

Rodrigo Costas
Thomas Franssen
Alfredo Yegros-Yegros

Layout

Andrea Reyes Elizondo
Suze van der Luijt-Jansen

The articles of this collection can be accessed at https://hdl.handle.net/1887/64521

ISBN: 978-90-9031204-0

23rd International Conference on Science and Technology Indicators (STI 2018)
"Science, Technology and Innovation indicators in transition"
12 - 14 September 2018 | Leiden, The Netherlands
#STI18LDN

A Bibliometric Framework for Identifying "Academic Chain"[1] ——A Case Study of 2014 Nobel Prize for Chemistry

Qianying Guo*, Jian Du** and Xiaoli Tang**

*guo.qianying@imicams.ac.cn
Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, 100005(China)

** windowsdujian@163.com; tang.xiaoli@imicams.ac.cn
Institute of Medical Information & Library, Chinese Academy of Medical Sciences, Beijing, 100005(China)

Introduction

Scientific research is an ever-evolving process, always with existing research results as the cornerstone of the new research results, as Ziman (2000) said, "Original ideas seldom come entirely 'out of the blue'. They are typically novel combinations of existing ideas". Academic inheritance is the internal logic of academic development and academic achievements form a chain based that, which is named academic chain (Shaohuai Liu, 2011). However, it's agreed that the emergence and development of a specific field depends on the few and extremely important documents, which can be arranged in chronological order and then form a story line. So the academic chain is a chain formed by articles which are essential to promote the development of the specific field as nodes and academic inheritance relationship as the link.

The research of academic chain can help to put the development process of the subject in order and then to promote its further development. There have been some achievements in the research of academic chain, such as summarizing the types of academic chain by the changes of the school specialization (Shaohuai Liu, 2011) or case analysis (Zhengfeng Li, 2012); characterizing the power of academic inheritance by Evaluative reference (Hairu Shang, Changgen Feng & Liang Sun, 2016) or the weights of traverse (Ge Song, 2017); identifying the seminal papers by Citation-Assisted Background (CAB)(Kostoff & Shlesinger, 2005); detecting the publications which are the historical roots (Bornmann & Marx, 2013; Marx, Bornmann & Barth, 2014), seminal research (Comins & Hussey, 2015) or research milestone (Comins & Leydesdorff, 2017) by cited references analysis or long-term references analysis. Reference relationship between academic papers is the main manifestation of the inheritance relationship between academic achievements, however, the papers which be cited for a long time or cited highly, is just be the results identified by one part of reference characteristics, and they would not be the all nodes of academic chain. Therefore, we combined the external features and content features of the citation relationship, then constructed a theoretical framework

[1] This work was supported by the National Natural Science Foundation of China (Grant No. 71603280) and CAMS Initiative for Innovative Medicine (CAMS-I2M-3-018).

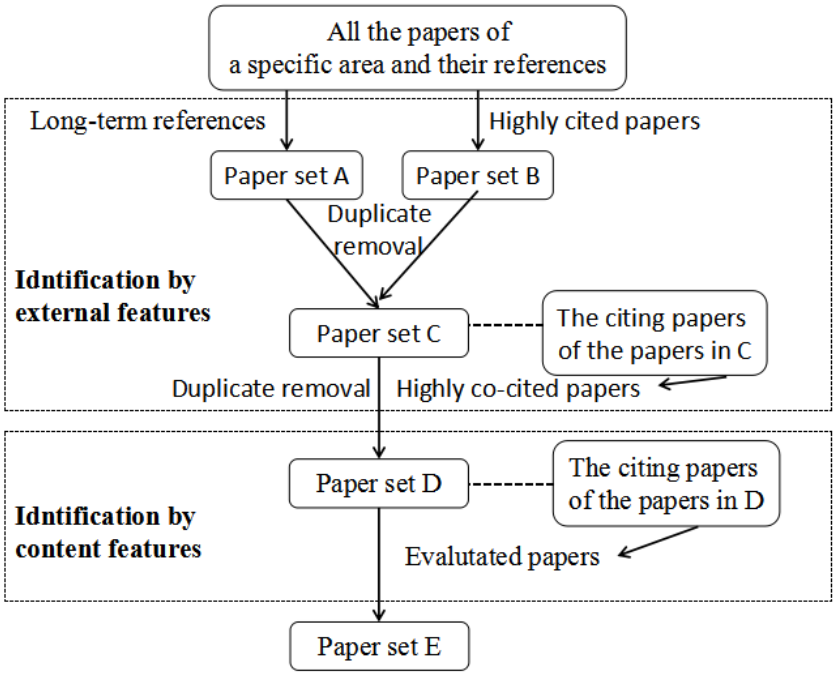
including the duration, scale, and content of the reference in this paper to identify the academic chain nodes.

Theoretical Basis and Identification Framework

The theoretical Basis of The Identification Framework

The reference relationship of academic papers mainly include the duration, scale, and content of reference. From the view of the duration of reference, "Only papers following the trajectory of a 'sticky knowledge claim' can be expected to have a sustained impact", which means the papers continue to be cited for more than ten years after publication (Baumgartner & Leydesdorff, 2014); in terms of the scale of reference, paper cited frequency is one of the current mainstream paper evaluation index, which can reflect the focus and recognition of peer scholars on their academic achievements to a certain extent, and there is complemented with the co-citation analysis considering the potential papers which are always co-mentioned with the pioneering academic papers but mighty be ignored by the hysteresis of the reference duration and cited frequency (Jian Du, 2017); in the matter of the content of reference, the symbolic evaluation about the paper's inheritance reflects the peer scholars' recognition on the inheritance to a certain extent (Hairu Shang, Changgen Feng & Liang Sun, 2016). Therefore, we tried to construct the identification framework including the reference duration, the cited frequency, the co-cited frequency and the evaluative reference to identify the academic chain nodes with the help of RPYS.exe, HistCite, Excel and other tools. The framework is shown in figure 1.

Figure 1: Theoretical identification framework of the academic chain



Description of Identification Framework

(1) Analysis of the reference duration. Identifying the long-term references by drawing the

reference publication year spectroscopy (RPYS) and the Multi-RPYS (Marx et al., 2014; Comins & Leydesdorff, 2016). We can find the references at peaks in the spectrum by RPYS, and then pick out the long-term references by Multi-RPYS. However, the RPYS i/o (Comins & Leydesdorff, 2016) just can handle the data whose publication years are between the year of 1900-1990, so we chose the RPYS.exe and Excel to achieve that. We put all the long-term references of the specific subject into the paper set A.

(2) Analysis of the cited frequency. Put the highly cited papers whose cited frequencies are in the top five by HistCite into the paper set B. Merge the paper set A and the paper set B and remove duplicate items, and then go back to the paper set C.

(3) Analysis of the co-cited frequency. Put the highly co-cited papers with each paper in the paper set C whose co-cited frequencies are row in the top five by HistCite into the paper set D, and then remove duplicate items.

(4) Analysis of the evaluative reference. According to "first", "broken", "breakthrough" and other iconic filtered comment words, if the citing papers whose cited frequencies are row in the top ten in its field by Web of Science are not referenced with such comment, we temporarily exclude the papers and eventually get the paper set E.

All papers in the paper set E are the nodes of academic chain identified by the framework, and then we can verify the correctness by comparing them with the actual nodes through coverage rate and accuracy rate.

Case study - a case study of the stimulated emission depletion (STED) imaging method of the 2014 Nobel Prize in chemistry

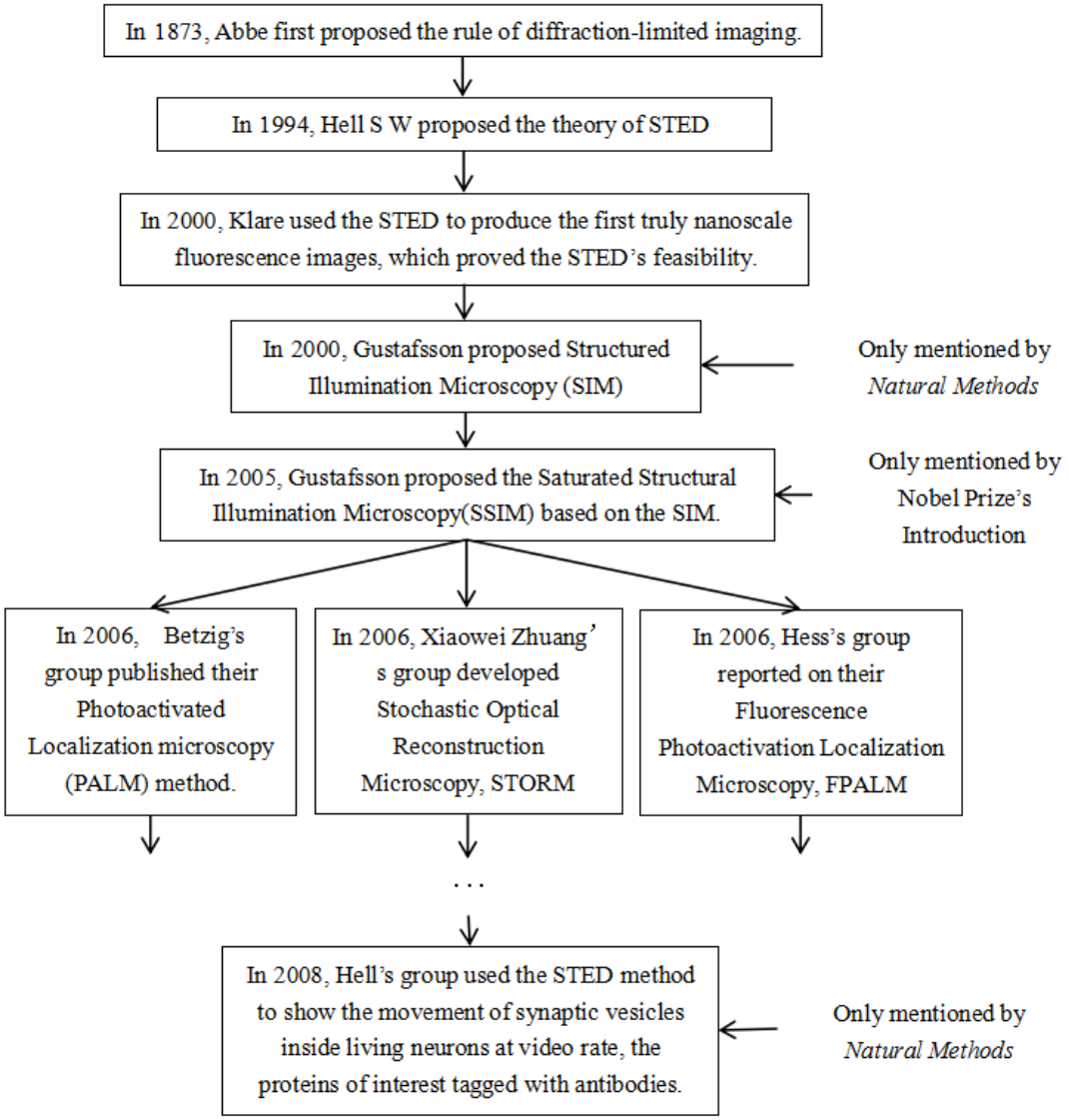
Topic introduction and data sources

Based on the 2014 Nobel Prize in chemistry, we arranged the previous life of the STED technology illustrated by the official introduction of the Nobel Prize and a review about the super-resolution microscopy published in *Natural Methods* in 2009 (Chi, 2009). Then we used the framework proposed in this paper to find the academic chain nodes. Finally, verify the feasibility of the identification framework by contrasting our results and the facts. Jian Du & Yishan Wu (2015) had used this case to identify the "sleeping beauty" and the princes who waken up it, and it has the larger feasibility in the long-term reference analysis and the co-citation analysis, so we chose this case.

Summarizing the history of STED technology in the introduction of Nobel Prize and the review of *Natural Methods* to be a reference standard as shown in figure 2.

We used the Web of Science core collection database to search for topics using ("Stimulated emission depletion" OR (STED microscopy)). The retrieval period was from 1970 to the time we retrieved (all time ranges covered by the database). The retrieval time was September 12, 2017. Finally, we obtained 944 related records. Using the above-mentioned identification framework of academic chain for analysis, the following results are obtained.

Figure 2: The development process of STED



(Arranged by *Scientific Background on the Nobel Prize in Chemistry 2014* & *Natural Methods*, 2008, 6(1):15-18)

Results and Analysis

As shown in Figure 3, the peak values of RPYS are in the year of 1873, 1959, 1994, 2000, 2006, 2008, and 2011. We can find the papers at each peak. It is worth noting that the peak is usually caused by a single paper, but not always that. A total of 10 peak papers were found here to draw their annual citation, as shown in Figure 4. The two papers of Richards (1959) and Vicidomini (2011) were excluded according to "continue to be cited more than 10 years after publication".

Figure 3: RPYS of the papers in the field of STED

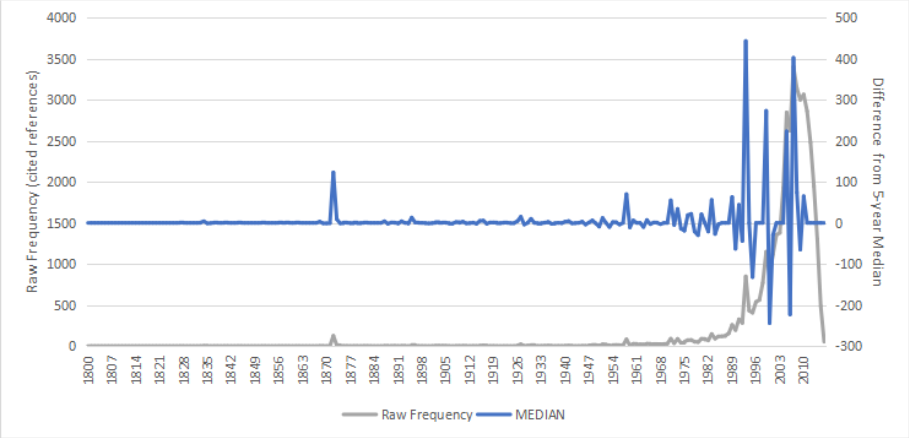
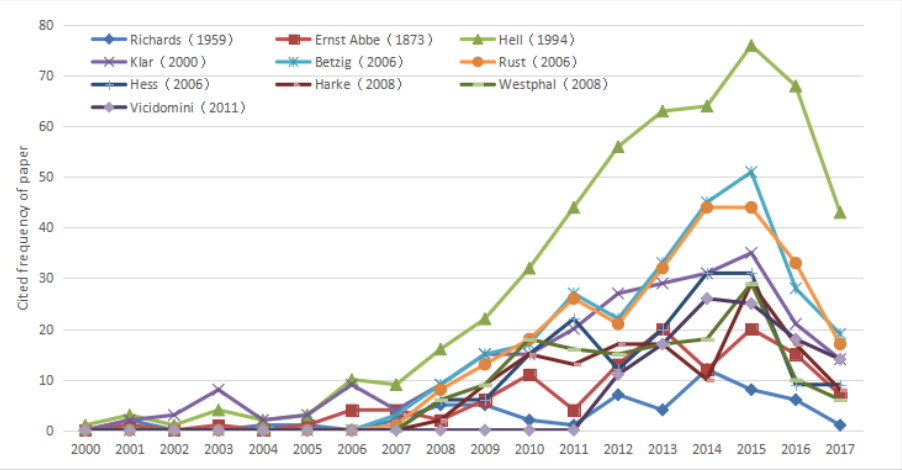


Figure 4: Citation curves of papers at peaks.



According to the "first", "broken", "breakthrough" and other iconic filtered comment words, the results above are filtered again and the final results are shown in table 1.

Table 1. The list of papers in paper set E.

No.	The first author (publication year)	Identification mark	The first author (publication year) of classical citing paper
1	Abbe E.(1873)	was originally recognized/ was first recognized by	Klar T A.(2001)
2	Hell S W.(1994)	opens up a new avenue	Eggeling, Christian(2009)
3	Klar T A.(2000)	the first ...technique	Fernandez-Suarez(2008)
4	Gustafsson M G.(2005)	has been broken by/ further improved	Ding J B.(2009)
5	Betzig E.(2006)	breakthrough	Toomre D(2010)
	Rust M J.(2006)		
	Hess S T.(2006)		
6	Westphal V.(2008)	First... succeeded by	Schermelleh L(2010)

As shown in Table 2, 8 papers were continuous cited for more than ten years, of which 7 were nodes of academic chain. The 4 of 5 papers most cited were duplicates and 1 paper was supplemented, but not the nodal paper. A total of 12 papers were obtained by the 5 most co-cited papers with each of the above 9 papers, and 3 papers were added, of which 2 were nodal papers. Using the three external features of long-term reference, highly cited reference and highly co-cited reference to identify the academic chain nodes, the total coverage rate is 100%, but the accuracy rate is 66.67%. The results identified by the highly cited index of the papers were either duplicated or not the academic chain nodes. 4 articles were excluded from the evaluative reference, and 8 articles were finally obtained. 8 articles all had academic inheritance significance, which increased the accuracy rate, but also reduced the coverage rate. The only missing result of the identification framework was the SIM technology proposed by Gustafsson in 2000. However, it must be pointed out that the paper has been included in the paper set D by the external features, and has been excluded after being identified by the content features. This shows that only selecting the citing papers with the cited frequency of the top ten may be insufficient.

Table 2. Identification results of academic chain theoretical framework.

The first author (publication year)	**Long-term references**	**highly cited papers**	**highly co-cited papers**	**Evaluated papers**
Abbe E.(1873)	$\surd^{\Delta}$	--	$\surd^{\Delta}$	$\surd^{\Delta}$
Hell S W.(1994)	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$
Klar T A.(2000)	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$
Gustafsson M G.(2005)	--	--	$\surd^{\Delta}$	$\surd^{\Delta}$
Betzig E.(2006)	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$
Rust M J (2006)	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$	$\surd^{\Delta}$
Hess S T.(2006)	$\surd^{\Delta}$	--	$\surd^{\Delta}$	$\surd^{\Delta}$
Westphal V (2008)	$\surd^{\Delta}$	--	$\surd^{\Delta}$	$\surd^{\Delta}$
Gustafsson M G.(2000)	--	--	$\surd^{\Delta}$	--
Hell S W (2007)	--	$\surd$	$\surd$	--
Harke (2008)	$\surd$	--	$\surd$	--
Willig K I (2007)	--	--	$\surd$	--
Cumulative coverage (%)	77.78	77.78	100	88.89
Cumulative accuracy (%)	87.5	77.78	75	100

Note：”$\surd$” represents the paper can be identified by the index;
”$\surd^{\Delta}$” represents the paper is the node of academic chain actually.

Conclusion

This paper proposes a theoretical framework for the identification of academic chain nodes, including the identification of external reference features of papers with long-term reference, highly cited reference and highly co-cited reference, supplemented by software tools such as

RPYS.exe, Excel and HistCite. The papers that have long-term influence or are cited by a large number of papers in the field are used as candidate papers, and in combination with the evaluation about the academic inheritance, which determines whether they are included in the final collection of papers.

The feasibility of this framework was verified by the history of STED technology which is reviewed by experts in the 2014 Nobel Prize and *Natural Methods* (Chi, 2009), and the result shows that the academic chain nodes were identified by this framework are basically consistent with the milestone events in the actual development process. At the same time, there are still some details that need further study:
(1) Whether it is reasonable to define the scope as the top five when selecting highly cited papers and highly co-cited papers.
(2) For content analysis of papers in paper set D, whether it is reasonable to define the citing papers of each paper in the paper set D as the top ten cited papers, and whether it is comprehensive for the selection of comment words. At the same time, we have not taken into account a series of non-standard reference, such as unmarked reference, indirect reference to non-original papers, etc., which inevitably result in some omissions.

In general, the theoretical framework for identifying the academic chain nodes proposed in this paper has certain scientificity and feasibility, and can identify milestone papers in the subject area. This is helpful to find the root papers in the field and understand the development of the field.

References List

Baumgartner S E, Leydesdorff L. Group-based Trajectory Modeling GBTM of Citations in Scholarly Literature: Dynamic Qualities of "Transient" and "Sticky Knowledge Claims"[M]. New York: John Wiley & Sons, Inc., 2014: 797-811.

Bornmann L, Marx W. The proposal of a broadening of perspective in evaluative bibliometrics by complementing the times cited with a cited reference analysis[J]. *Journal of Informetrics*, 2013, 7(1): 84-88.

Chi K R. Super-resolution microscopy: breaking the limits[J]. *Nature Methods*, 2009, 6(1): 15-18.

Comins J A, Hussey T W. Detecting seminal research contributions to the development and use of the global positioning system by reference publication year spectroscopy[J]. *Scientometrics*, 2015, 104(2): 575-580.

Comins J A, Leydesdorff L. RPYS i/o: software demonstration of a web-based tool for the historiography and visualization of citation[J]. *Scientometrics*. 2016, 107(3): 1509-1517.

Comins J A, Leydesdorff L. Citation algorithms for identifying research milestones driving biomedical innovation[J]. *Scientometrics*, 2017, 110(3): 1495-1504.

Ge Song. Building and Demonstrating "Creativity-Influence-Transmission" Research Evaluation Dimensions [In Chinese][J]. *Library and Information Service*. 2017(10): 5-13.

Hairu Shang, Changgen Feng, Liang Sun. Evaluation of academic papers with academic influence —Proposing two new indicators of academic inheritance effect and long-term citation [In Chinese] [J]. *Chinese Science Bulletin*, 2016(26): 2853-2860.

Jian Du. A study on systematic identification of "sleeping beauty" publications and their awaking mechanisms [In Chinese][D]. Nanjing: Nanjing University, 2017.

Jian Du, Yishan Wu. Identifying "Sleeping Beauties" and the Princes Who Woke Up Sleeping Beauies in Terms of Citation Speed Measures: A Case Study of the Key Publication by Stefan W. Hell [In Chinese] [J]. *Journal of the China Society for Scientific and Technical Information*, 2015(5): 508-521.

Kostoff R N, Shlesinger M F. CAB: Citation-Assisted Background[J]. *Scientometrics*, 2005, 62(2): 199-212.

Marx W, Bornmann L, Barth A, et al. Detecting the historical roots of Research Fields by Reference Publication Year Spectroscopy (RPYS)[J]. *Journal of the Association for Information Science & Technology*, 2014, 65(4): 751-764.

Shaohuai Liu. Academic chain:Objective existence of academic relations form [In Chinese].[J]. *Thinking*, 2011(1): 1-3.

Zhengfeng Li. The Inheritance and Innovation of the Chinese Scientists'
Academic Thoughts: Concepts， Features and Methods [In Chinese][J]. *Nanjing Journal of Social Sciences*. 2012(04): 1-8.

Ziman, J. (2000). Real science: What it is, and what it means. Cambridge, UK: Cambridge University Press.