

# Comparative Study of Trace Metrics between Bibliometrics and Patentometrics

Fred Y. Ye<sup>1</sup>, Mu-Hsuan Huang<sup>2</sup> & Dar-Zen Chen<sup>3†</sup>

<sup>1</sup>School of Information Management, Nanjing University, Nanjing 210023, China

<sup>2</sup>Department of Library and Information Science, Taiwan University, Taipei 10617, China

<sup>3</sup>Department of Mechanical Engineering and Institute of Industrial Engineering, Taiwan University, Taipei 10617, China

Citation: Fred Y. Ye, Mu-Hsuan Huang & Dar-Zen Chen (2016). Comparative Study of Trace Metrics between Bibliometrics and Patentometrics.

Received: Feb. 20, 2016

Revised: Apr. 15, 2016

Accepted: May 12, 2016

## Abstract

**Purpose:** To comprehensively evaluate the overall performance of a group or an individual in both bibliometrics and patentometrics.

**Design/methodology/approach:** Trace metrics were applied to the top 30 universities in the 2014 Academic Ranking of World Universities (ARWU)— computer sciences, the top 30 ESI highly cited papers in the computer sciences field in 2014, as well as the top 30 assignees and the top 30 most cited patents in the National Bureau of Economic Research (NBER) computer hardware and software category.

**Findings:** We found that, by applying trace metrics, the research or marketing impact efficiency, at both group and individual levels, was clearly observed. Furthermore, trace metrics were more sensitive to the different publication-citation distributions than the average citation and *h*-index were.

**Research limitations:** Trace metrics considered publications with zero citations as negative contributions. One should clarify how he/she evaluates a zero-citation paper or patent before applying trace metrics.

**Practical implications:** Decision makers could regularly examine the performance of their university/company by applying trace metrics and adjust their policies accordingly.

**Originality/value:** Trace metrics could be applied both in bibliometrics and patentometrics and provide a comprehensive view. Moreover, the high sensitivity and unique impact efficiency view provided by trace metrics can facilitate decision makers in examining and adjusting their policies.

**Keywords** Performance matrix; Trace metrics; *h*-index; *h*-core; I3; Bibliometrics; Patentometrics



JDIS  
Journal of Data and  
Information Science  
Vol. 1 No. 2, 2016  
pp 13–31

DOI: 10.20309/jdis.201611

<http://www.jdis.org>

<sup>†</sup> Corresponding author: Dar-Zen Chen (E-mail: dzchen@ntu.edu.tw).

## 1 Introduction

Performance and efficiency evaluation is an essential but challenging task for managers in fields ranging from science to business. Therefore, in bibliometrics, several citation indicators, including intuitive indicators such as total and average citation counts and extended indicators such as the impact factor (IF) (Garfield, 1972) and  $h$ -index (Hirsch, 2005) have been designed to evaluate the academic performance of a university or researcher or other units. Narin, Noma, and Perry (1987) first used patents as an indicator for measuring the technological strength of a corporation. Although the aforementioned indicators have been widely applied in literature and bibliographic databases, they have some limitations. For example, the skewness of citation distributions is ignored in the citation counts and IF (Leydesdorff & Bornmann, 2011) and the  $h$ -index are somewhat inconsistent (Waltman & van Eck, 2012) and insensitive (Bornheim et al., 2008; Egghe, 2006; Kuan, Huang, & Chen, 2011).

Within a researcher's publication set, the rank distribution of citations should theoretically be a curve. The publication set is likely to include certain highly cited papers and many scarcely cited papers (Bornmann, Mutz, & Daniel, 2010), but the  $h$ -index reflects only the  $h \times h$  area. Moreover, individual researchers with a dissimilar citation distribution may have the same  $h$ -index value (Bornmann et al., 2010; García-Pérez, 2009). The rank-citation curve overcomes the limitations of the  $h$ -index by representing a researcher's performance over a particular period (Kuan et al., 2011). The tapered  $h$ -index summarizes the impact of every citation in the citation curve by weighting the citations on the basis of the Durfee square (Anderson, Hankin, & Killworth, 2008). García-Pérez (2009) proposed an iterative view of the  $h$ -index in which the rank-citation curve is divided into several  $h$ -indices to demonstrate the differences in the citation distribution among individual researchers (García-Pérez, 2009). Bornmann et al. (2010) proposed three areas under the rank-citation curve: an area that has citations lower than the  $h$ -index ( $h_2$  lower;  $t$ -area in Figure 1), a square area captured by the  $h$ -index ( $h_2$  center;  $h$ -area in Figure 1), and an area in which citations exceed the  $h$ -index ( $h_2$  upper,  $e$ -area in Figure 1). Leydesdorff and Bornmann (2011) proposed using integrated impact indicators (I3s) instead of the IF for evaluating academic performance (Leydesdorff & Bornmann, 2011).

According to the definition of I3 (Leydesdorff & Bornmann (2011), an I3-type indicator can be formalized as

$$I3(i) = \sum_{i=1}^C f(X_i) \cdot X_i . \quad (1)$$

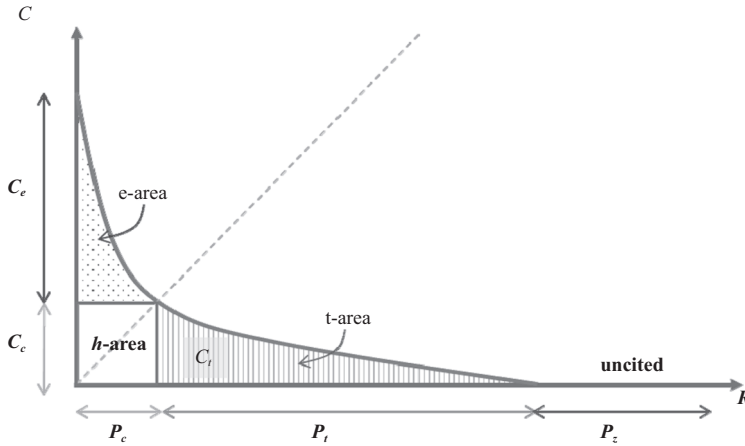


Figure 1. Rank-citation curve with information on the number of publications. The area under the rank-citation curve is divided into four sections: the  $h$ -area, based on the  $h$ -index; the  $e$ -area, containing the excess citations of the first  $h$  papers to the  $h$ -area; the  $t$ -area, containing citations of the papers that has lower citations than  $h$ , but still representing a contribution; and the uncited area.

where  $X_i$  indicates the percentile ranks and  $f(X_i)$  indicates the frequencies of the ranks,  $i$  in  $[1, C]$  indicates the percentile rank classes. The number  $C$  is the total classes that the measures  $X_i$  are divided into, each with a scoring function  $f(X_i)$  or weight ( $w_i$ ). Therefore, the  $I3$ -type indicator can also be written as

$$I3(i) = \sum_i w_i X_i. \quad (2)$$

Similar to  $I3$ , if a weighted  $I3$ -type measure corresponding to publications and citations in the  $h$ -core and  $h$ -tail framework is proposed (c.f. Figure 1), an  $I3$ -like publication indicator ( $I3X$ ) and an  $I3$ -like citation indicator ( $I3Y$ ) can be defined on the basis of the three classes as follows:

$$I3X = x_c P_c + x_t P_t + x_z P_z = \frac{P_c}{P_c + P_t + P_z} \cdot P_c + \frac{P_t}{P_c + P_t + P_z} \cdot P_t + \frac{P_z}{P_c + P_t + P_z} \cdot P_z, \quad (3)$$

$$I3Y = y_c C_c + y_t C_t + y_e C_e = \frac{C_c}{C_c + C_t + C_e} \cdot C_c + \frac{C_t}{C_c + C_t + C_e} \cdot C_t + \frac{C_e}{C_c + C_t + C_e} \cdot C_e, \quad (4)$$

in which the weighting scores for  $P_c, P_t, P_z, C_c, C_t$ , and  $C_e$  become  $x_c = P_c / (P_c + P_t + P_z)$ ,  $x_t = P_t / (P_c + P_t + P_z)$ ,  $x_z = P_z / (P_c + P_t + P_z)$ ,  $y_c = C_c / (C_c + C_t + C_e)$ ,  $y_t = C_t / (C_c + C_t + C_e)$ , and  $y_e = C_e / (C_c + C_t + C_e)$ , respectively.



**Research Paper**

The publication vector  $X$  and citation vector  $Y$  can then be defined, and  $Z$  can be introduced as follows:

$$X = (X_1, X_2, X_3) = (x_c P_c, x_t P_t, x_z P_z), \quad (5)$$

$$Y = (Y_1, Y_2, Y_3) = (y_c C_c, y_t C_t, y_e C_e), \quad (6)$$

$$Z = (Z_1, Z_2, Z_3) = (Y_1 - X_1, Y_2 - X_2, Y_3 - X_3). \quad (7)$$

When the  $h$ -index is combined with  $I3$ ,  $3 \times 3$  performance matrices  $V_1 = (Y, X, Z)^T$  and  $V_2 = (X, Y, Z)^T$  can be constructed. Accordingly, if an indicator is required for comparing or ranking scholarly individuals or groups, the traces of performance matrices that provide scalars that summarize academic performance, such as  $T_1 = \text{Tr}(V_1) = Y_1 + X_2 + Z_3$  and  $T_2 = \text{Tr}(V_2) = X_1 + Y_2 + Z_3$ , can be computed. Therefore, multivariate information in the citation curve can be expressed in single measures.

Because trace metrics summarize all the information in the citation curve, they can be applied for measuring the overall performance of a university, assignee, paper, or patent. The remainder of the paper is organized as follows. Section 2 provides a detailed explanation of how trace metrics were calculated and how data were chosen. Section 3 presents the results. Finally, Section 4 presents the discussions and conclusions.

## 2 Methodology

### 2.1 Method

We extended the performance matrix proposed by Ye and Leydesdorff (2014) to a primary matrix  $V_1$ , a secondary matrix  $V_2$ , and a submatrix  $SV$  (Huang et al., 2015), which consider the overall effects of citation distribution and publication distribution.

$$V_1 = \begin{pmatrix} Y_1 & Y_2 & Y_3 \\ X_1 & X_2 & X_3 \\ Z_1 & Z_2 & Z_3 \end{pmatrix} = \begin{pmatrix} Y \\ X \\ Z \end{pmatrix}, \quad (8)$$

$$V_2 = \begin{pmatrix} X_1 & X_2 & X_3 \\ Y_1 & Y_2 & Y_3 \\ Z_1 & Z_2 & Z_3 \end{pmatrix} = \begin{pmatrix} X \\ Y \\ Z \end{pmatrix}, \quad (9)$$

$$SV = \begin{pmatrix} Y_h & Y_2 \\ X_1 & X_2 \end{pmatrix}, \quad (10)$$



where  $X_i = P_i \frac{P_i}{P}$  is an I3-type score of publications and  $Y_i = C_i \frac{C_i}{C}$  is an I3-type

score of citations. For  $V_1$  and  $V_2$ ,  $X_1 = P_c \frac{P_c}{P}$ ,  $X_2 = P_t \frac{P_t}{P}$ , and  $X_3 = P_z \frac{P_z}{P}$ , whereas

$Y_1 = C_c \frac{C_c}{C}$ ,  $Y_2 = C_t \frac{C_t}{C}$ , and  $Y_3 = C_e \frac{C_e}{C}$ . For  $SV$ ,  $Y_h = C_h \frac{C_h}{C}$ ,

where  $C_c$  is the number of citations in the  $h$ -core-area, equaling  $h^2$ ;

$C_t$  is the number of citations in the  $t$ -area;

$C_e$  is the number of citations in the  $e$ -area;

$C_h$  is the number of citations of the  $h$ -area,  $C_h = C_c + C_e$ ;

$C$  is the total number of citations, equaling  $C_c + C_t + C_e$ ;

$P_c$  is the number of papers in the  $h$ -area, equaling  $h$ ;

$P_t$  is the number of papers in the  $t$ -area;

$P_z$  is the number of papers having zero citations; and

$P$  is the total number of papers, equaling  $P_c + P_t + P_z$ .

The vectors  $X = (X_1, X_2, X_3)$  and  $Y = (Y_1, Y_2, Y_3)$  are publication and citation vectors, respectively.

The three traces of matrices  $V_1$ ,  $V_2$ , and  $SV$  can then be used to obtain indicators  $T_1$ ,  $T_2$ , and  $ST$  as follows:

$$T_1 = \text{Tr}(V_1) = Y_1 + X_2 + Z_3 = \frac{C_c^2}{C} + \frac{P_t^2}{P} + \left(\frac{C_e^2}{C} - \frac{P_z^2}{P}\right), \quad (11)$$

$$T_2 = \text{Tr}(V_2) = X_1 + Y_2 + Z_3 = \frac{P_c^2}{P} + \frac{C_t^2}{C} + \left(\frac{C_e^2}{C} - \frac{P_z^2}{P}\right), \quad (12)$$

$$ST = \text{Tr}(SV) = Y_h + X_2 = \frac{C_h^2}{C} + \frac{P_t^2}{P}. \quad (13)$$

Both  $T_1$  and  $T_2$  summarize the representative information distributed over the  $e$ -,  $h$ -,  $t$ -, and uncited areas in the rank-citation graph.

For a demonstration of trace metrics, we applied traces  $T_1$ ,  $T_2$ , and  $ST$  to both bibliometrics and patentometrics to investigate the performance of an institution (e.g. a university or a company) and that of a single document (e.g. a paper or a patent). The traces in group level can be called academic traces (for universities or scientists) or assignee traces (for companies or assignees) and those in individual level can be called impact traces (for papers) or patent traces (for patents).

In this research, we used full counts to assign credits of publications to organizations. Although some might debate that using full counts in bibliometrics



would magnify the actual number of publications, full counting is the most intuitive and currently most widely-used counting method in bibliometrics. From the perspective of patents, only few patents have more than one assignee. Zheng et al. (2013) studied the influence of counting methods in patentometrics and found that the difference among different counting methods is slight. In this preliminary research of applying trace metrics in bibliometrics and patentometrics, we chose to compare the trace metric performance of universities and companies using a full counting method and to leave the author contribution-credit issue to future work.

## 2.2 Data

We used the traces  $T_1$ ,  $T_2$ , and  $ST$  on both bibliometrics and patentometrics. For an informetric test, we applied traces to investigate the performance of the top 30 universities with respect to the computer sciences according to the 2014 Academic Ranking of World Universities (ARWU) – computer sciences. We also applied traces to the top 30 most cited papers from Essential Science Indicators (ESI) Highly Cited Papers – computer sciences, published in March 2015. The five year (i.e. from 2010/01/01 to 2014/12/31) bibliographic data of the top universities and the highly cited papers were collected from the Web of Science database updated on 2015/04/08, which means the citation counts were accumulated from 2010/01/01 to 2015/04/08. In order to compare the trace performances among the 30 universities in the computer sciences, we applied the ESI journal list to confine the bibliometric data we analyzed to the field of computer sciences.

For a patentometric test, we selected the top 30 assignees who owned the most patents in the National Bureau of Economic Research (NBER) computer hardware and software category that were issued from 2010/01/01 to 2014/12/31. Similar to the procedure used for the bibliometric test, we selected the top 30 most cited US patents in the NBER computer hardware and software category that were issued from 2010/01/01 to 2014/12/31. All patent data were obtained from the United States Patent and Trademark Office database.

The datasets covered group level (universities and companies) and individual level (paper and patent, individually as a single publication). For calculating the traces of a single document (a highly cited paper or patent), we followed Schubert's (2009) method to construct a rank-citation graph of the single document by determining the number and citations of citing documents (i.e. documents that cite the document under consideration). Therefore, the  $h$ -index of a single document could be determined.



### 3 Results

#### 3.1 Comparison at Group Level: Academic and Assignee Traces

Applying trace metrics to a university enables assessing its academic performance. We call such metrics academic traces.

Figure 2 shows the values of academic traces  $T_1$  (solid blue line with squares),  $T_2$  (solid orange line with triangles), and  $ST$  (solid green line with circles); this figure also shows the typical academic indicators of average citations per paper ( $C/P$ , brown dashed line with x) and the  $h$ -index (gray bar plot) of the top 30 computer science universities in the ARWU 2014 subject ranking. From left to right, the universities are listed in descending order according to total citations. Academic traces  $T_1$ ,  $T_2$ , and  $ST$  share the same scale and are expressed along the left vertical axis, whereas  $C/P$  and the  $h$ -index were much lower than  $T_1$ ,  $T_2$ , and  $ST$ , and they are expressed along the right vertical axis. Generally,  $T_1$ ,  $T_2$ ,  $ST$ , and  $h$  followed this descending trend, except for Tsinghua University and Carnegie Mellon University (CMU), which exhibited a rise in  $T_2$  and a drop in the  $h$ -index. However, two universities, Taiwan University and Israel Institute of Technology (Technion), showed a drop in the  $h$ -index, but they did not show a rise in  $T_2$ . We carefully examined these four datasets and determined that, compared with other universities, Tsinghua University and CMU published more papers having numbers of citations that were lower than the  $h$ -index but higher than 0 (i.e. a higher  $P_i$ , resulting in a higher  $C_i$  and causing a square effect on  $T_2$ ). These  $h$ -drop- $T_2$ -rise universities were compared with  $T_2$ -drop- $ST$ -rise universities such as the University of California, Berkeley, University of California, San Diego, University of Toronto, University of Michigan, and California Institute of Technology (CalTech), which had fewer total papers  $P$  and thus a higher  $P_i^2/P$  and  $P_z^2/P$ . The  $T_2$ -drop- $ST$ -rise universities can be explained as being low-publication but high-citation universities, which can also be proven by their above-average  $C/P$  values.

Table 1 shows Pearson (bottom left part of the table, with no background) and Spearman (top right part of the table, with a gray background) correlation coefficients among  $T_1$ ,  $T_2$ ,  $ST$ ,  $C/P$ , and the  $h$ -index. The three trace metrics can be divided into two groups: The first group contained  $T_1$  and  $ST$ , which had high correlation coefficients with the commonly used bibliometric indicators  $C/P$  and  $h$ , whereas the second group contained  $T_2$ , which had a low correlation coefficient with the average indicator  $C/P$ , but it was still highly correlated with  $h$ . We found that, although both  $T_2$  and  $C/P$  are highly correlated with  $T_1$ ,  $ST$  and  $h$ , they do not show a correlation with each other. This means that  $T_2$  and  $C/P$  may provide us with important information as  $T_1$ ,  $ST$  and  $h$  do, but from different perspectives.



Research Paper

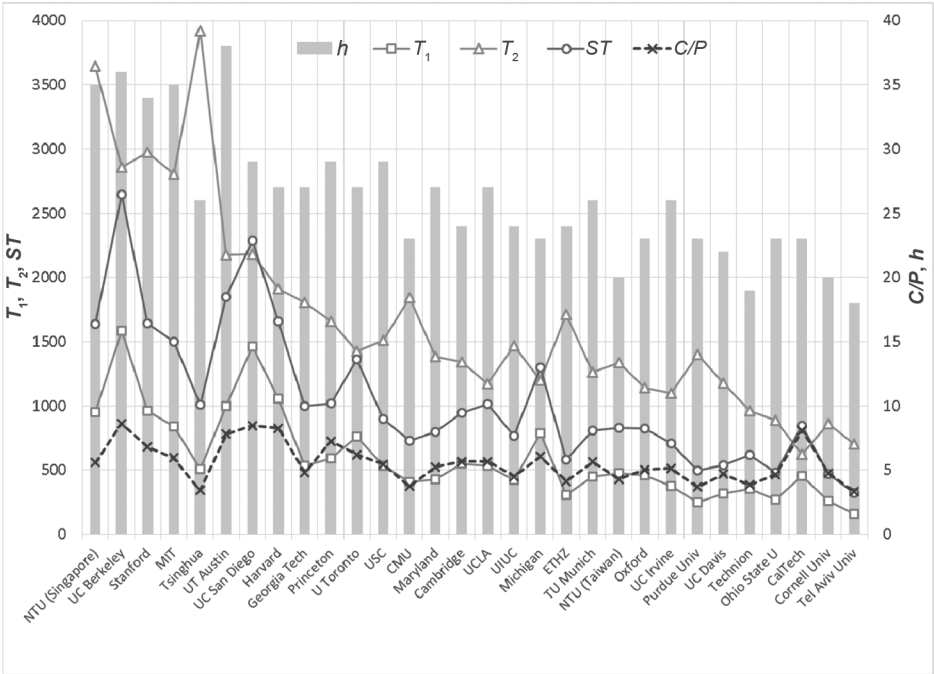


Figure 2. Academic traces  $T_1$ ,  $T_2$ , and  $ST$ ; citations per paper ( $C/P$ ), and  $h$ -index for the top 30 universities in computer sciences (2010–2014).

Table 1. Spearman and Pearson correlation coefficients among the  $C/P$ ,  $h$ ,  $T_1$ ,  $T_2$ , and  $ST$  of the top 30 universities.

Pearson \ Spearman		$T_1$	$T_2$	$ST$	$C/P$	$h$
		$T_1$	$T_2$	$ST$	$C/P$	$h$
	$T_1$	1	0.581*	0.988*	0.730*	0.721*
	$T_2$	0.593*	1	0.598*	0.065	0.755*
	$ST$	0.994*	0.624*	1	0.715*	0.749*
	$C/P$	0.766*	0.104	0.752*	1	0.460*
	$h$	0.751*	0.745*	0.790*	0.519*	1

Note. \* Significant correlation at 0.01 level.



Several bibliometric indicators are used in patentometrics for estimating the performance of patents. Similar to the procedures performed for bibliometrics, a company can be evaluated according to the performance of its patents. When trace metrics are applied to a group level of patent, they are called assignee traces.

Figure 3 illustrates the values of assignee traces  $T_1$  (blue solid line with squares),  $T_2$  (orange solid line with triangles), and  $ST$  (green solid line with circles); this figure



also shows the commonly used bibliometric indicators of average citations  $C/P$  (brown dashed line with x) and the  $h$ -index (gray bar plot) of the top 30 assignees in the NBER computer hardware and software category. The assignees are listed from left to right in descending order according to the total citations of their patents.  $T_1$ ,  $T_2$ , and  $ST$  share the same scale and are expressed along the left vertical axis, whereas  $C/P$  and the  $h$ -index were much lower than  $T_1$ ,  $T_2$  and  $ST$ , and are expressed along the right vertical axis. In general, all indicators followed a descending trend, except for IBM, which had the most citations and a relatively high  $h$ -index but a considerably negative  $T_1$  value. The reason for the drop in  $T_1$  is that IBM has many zero-citation patents, leading to a considerably large value for  $\frac{P_z^2}{P}$ , subtracted from a considerably lower value  $\frac{C_e^2}{C}$ . In addition, software companies such as Microsoft, Google, Oracle, Amazon, Yahoo, and Digimarc achieved more satisfactory trace performance levels than hardware companies such as IBM, Apple, Sony, HP, and SAP did.

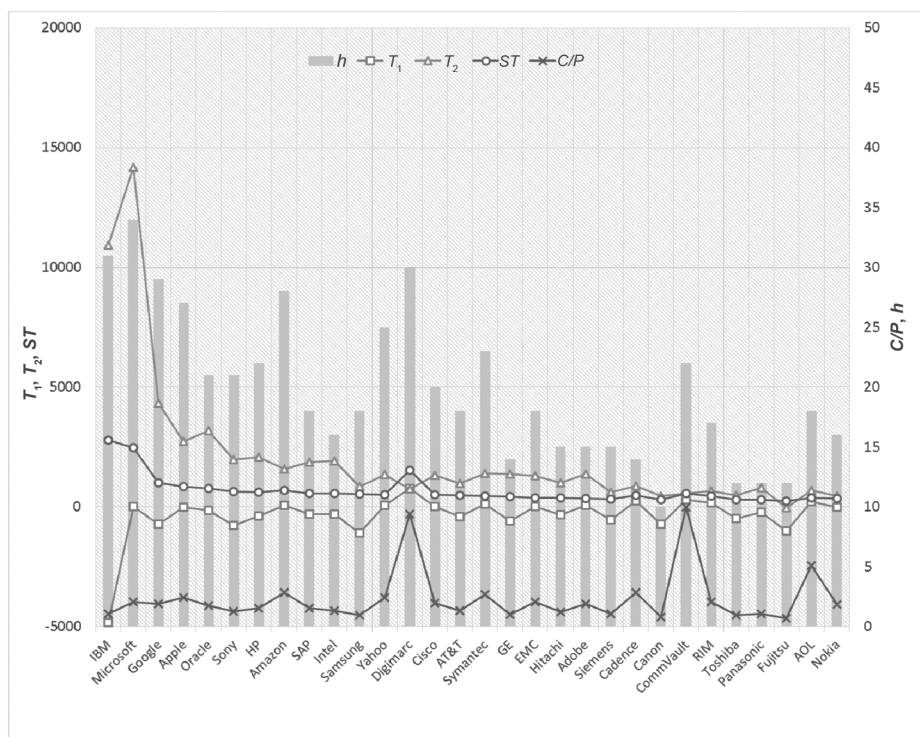


Figure 3. Values of  $T_1$ ,  $T_2$ ,  $ST$ ,  $C/P$ , and  $h$ -index for the top 30 assignees (2010–2014).

## Research Paper

Table 2 shows Pearson (bottom left part of the table, with no background) and Spearman (top right part of the table, with a gray background) correlation coefficients among  $T_1$ ,  $T_2$ ,  $ST$ ,  $C/P$ , and the  $h$ -index. In Pearson's correlation analysis,  $T_1$  negatively correlated with  $T_2$ ,  $ST$ , and the  $h$ -index (Table 3). Compared with papers, most patents had relatively low citation values and thus, a low  $h$ -index and  $C_e$  and a high  $P_z$ , leading to a low  $T_1$ ; hence, the trend of  $T_1$  was different from that of  $T_2$ ,  $ST$ , and the  $h$ -index.

Table 2. Spearman and Pearson correlation coefficients among the  $C/P$ ,  $h$ ,  $T_1$ ,  $T_2$ , and  $ST$  of the top 30 assignees.

	Spearman	$T_1$	$T_2$	$ST$	$C/P$	$h$
Pearson						
$T_1$		1	-0.248	0.093	0.932*	0.275
$T_2$		-0.507*	1	0.655*	-0.094	0.519*
$ST$		-0.482*	0.881*	1	0.275	0.799*
$C/P$		0.303	-0.099	0.177	1	0.488*
$h$		-0.151	0.646*	0.750*	-0.121	1

Note. \* Significant correlation at 0.01 level.

At the group level, we observed that for both universities and companies, the difference between their average citation and  $h$ -index values was small. The average citation value for the top 30 universities was approximately 5, whereas that for the top 30 companies was approximately 2. The  $h$ -index ranged from 15 to 40 for the universities, and it extended from 10 to 35 for the companies. The differences in trace metrics between universities and between companies were more significant. Most trace metrics varied from 0 to 2000 for the top 30 universities, and stretched from -1000 to 1500 for the top 30 companies. We considered zero citation as a negative contribution, and there were more zero-citation patents than zero-citation papers. Therefore, numerous companies had a negative  $T_1$  value. This negative value can be considered a warning, rather than being perceived as indicating no market value; accordingly, patents' potential market value should be investigated. Compared with an acceptable negative trace metric value in patentometrics, a negative trace metric value in bibliometrics indicates poor efficiency in conducting crucial research. Therefore, if a university receives a negative trace metric value, it should examine its research projects and consider adjusting them.

We determined that in contrast to the patentometric indicators, all bibliometric indicators showed significant correlations. This discrepancy means that bibliometric indicators as well as traces are generally applicable and that other factors such as market elements must be considered in patentometric indicators to ensure their applicability.



### 3.2 Comparison at the Individual Level: Impact and Patent Traces

In addition to the universities, the trace metrics were applied to a single paper to evaluate its impact. We called these metrics impact traces.

Figure 4 shows the values of impact traces  $T_1$  (solid blue line with squares),  $T_2$  (solid orange line with triangles), and  $ST$  (solid green line with circles) as well as the commonly used academic indicators of average citations  $C/P$  (brown dashed line with x) and the  $h$ -index (gray bar plot) of the top 30 most cited computer science papers according to ESI data obtained in March 2015. Table A1 lists detailed information on these papers. The most cited papers were named according to their rank in citation; that is, the most cited paper was named P1, and the second most cited paper was named P2. Impact traces  $T_1$ ,  $T_2$ , and  $ST$  share the same scale and are expressed along the left vertical axis, whereas  $C/P$  and the  $h$ -index were much lower than  $T_1$ ,  $T_2$ , and  $ST$ , and are expressed along the right vertical axis. In general, all trace metrics followed a descending trend from left to right, except for P5, which exhibited a rise in  $T_1$  and  $ST$ , and P3, P4, and P6, which exhibited a drop in  $T_1$  and  $ST$ . These results may be attributable to P5 having a relatively low number of citing

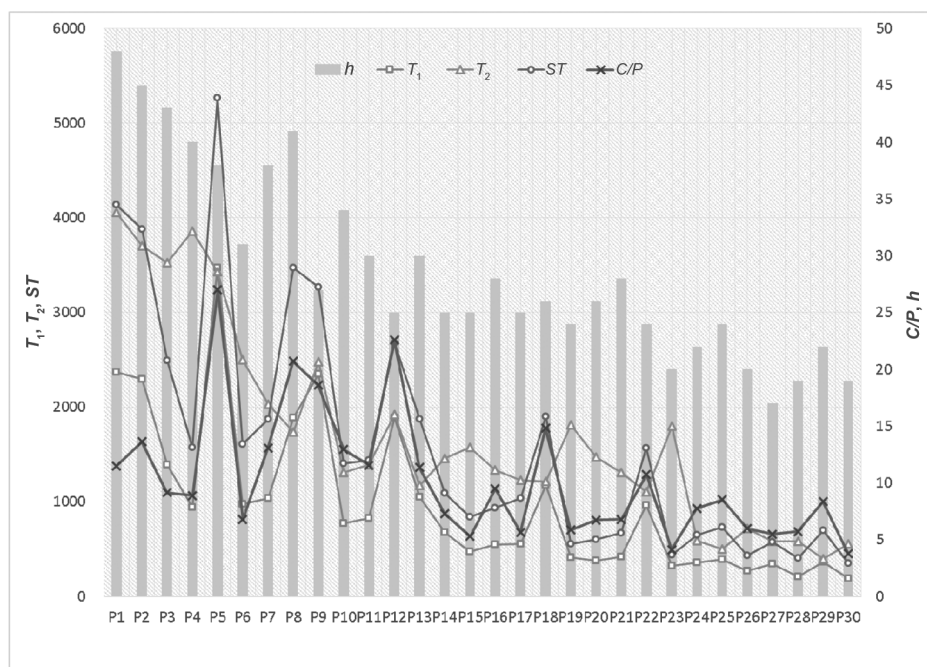


Figure 4. Impact traces  $T_1$ ,  $T_2$ , and  $ST$ ; citations per paper, and  $h$ -index for the top 30 highly cited computer science papers (2010–2014).

## Research Paper

papers with zero citations and P3, P4, and P6 having a relatively high number of citing papers with zero citations. Compared with academic traces, impact traces had similar values but were less consistent among  $T_1$ ,  $T_2$ , and  $ST$ .

Table 3 lists Pearson (bottom left part of the table, without a background) and Spearman (top right part of the table, with a gray background) correlation coefficients among  $T_1$ ,  $T_2$ ,  $ST$ ,  $C/P$ , and the  $h$ -index. For highly cited papers, impact traces were highly correlated with average citations and the  $h$ -index.

Table 3. Spearman and Pearson correlation coefficients among the  $C/P$ ,  $h$ ,  $T_1$ ,  $T_2$ , and  $ST$  of the top 30 highly cited papers.

	Spearman	$T_1$	$T_2$	$ST$	$C/P$	$h$
Pearson						
$T_1$		1	0.868*	0.994*	0.926*	0.922*
$T_2$		0.820*	1	0.838*	0.746*	0.907*
$ST$		0.992*	0.828*	1	0.926*	0.915*
$C/P$		0.901*	0.627*	0.880*	1	0.867*
$h$		0.793*	0.897*	0.827*	0.706*	1

Note. \* Significant correlation at 0.01 level.

Similar to our previous bibliometric analysis, the impact of a single patent was studied using trace metrics (subsequently denoted as patent traces).

Figure 5 illustrates the values of patent traces  $T_1$  (solid blue line with squares),  $T_2$  (solid orange line with triangles), and  $ST$  (solid green line with circles) as well as the commonly used indicators of average citations  $C/P$  (brown dashed line with x) and the  $h$ -index (gray bar plot) of the top 30 most cited patents in the NBER computer hardware and software category. Table A2 lists detailed information on these patents. Patent traces  $T_1$ ,  $T_2$ , and  $ST$  share the same scale and are expressed along the left vertical axis, whereas  $C/P$  and the  $h$ -index were much lower than  $T_1$ ,  $T_2$ , and  $ST$  and are expressed along the right vertical axis. These most cited patents are listed in descending order from left to right according to total citations. The top seven most cited patents exhibited a relatively low  $T_2$  value (Figure 5), and thus, a low Pearson correlation coefficient was observed between  $T_2$  and other indicators (Table 4). All the top seven most cited patents had a relatively high  $h$ -index, possibly indicating centrality in the  $h$ -core and thus a low  $C_e$  value. In contrast to the assignee traces, marked differences existed among patent traces  $T_1$ ,  $ST$ , and  $T_2$ . After carefully examining these patents, we observed that the hardware patents P1–P7 had a relatively high  $h$ -index value and thus, exhibited considerable differences among  $T_1$  and  $ST$ , which were dominated by  $h^4$ , and  $T_2$ , which was proportional to  $h^2$ .

Table 4 lists Pearson (bottom left part of the table, with no background) and Spearman (top right part of the table, with a gray background) correlation coefficients



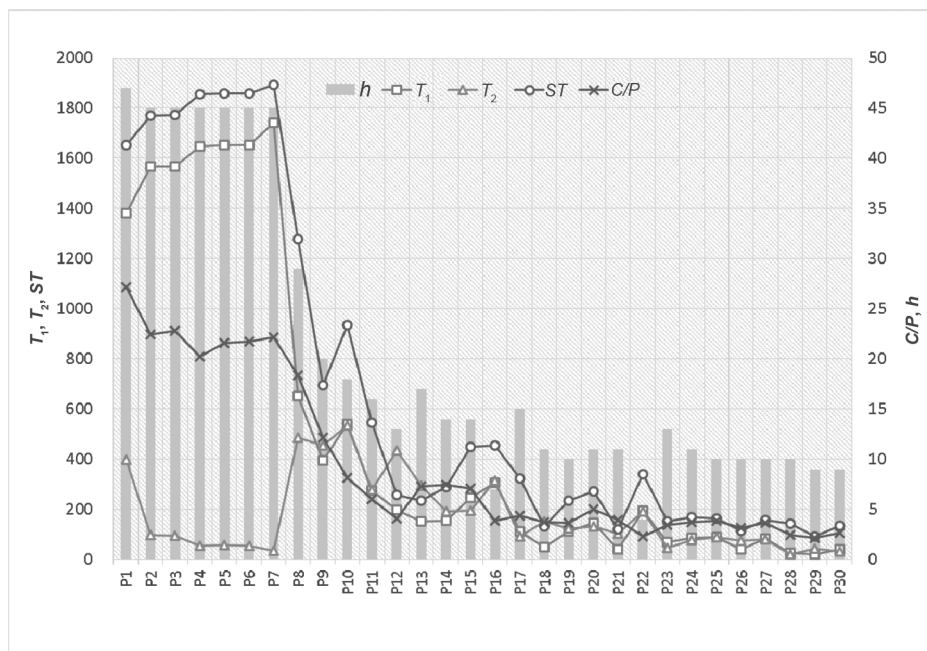


Figure 5. Values of  $T_1$ ,  $T_2$ ,  $ST$ ,  $C/P$ , and the  $h$ -index for the top 30 highly cited patents (2010–2014).

among  $T_1$ ,  $T_2$ ,  $ST$ ,  $C/P$ , and the  $h$ -index. Most of the indicators demonstrated a satisfactory correlation coefficient with other indicators, except for the Pearson correlation coefficient of  $T_2$ .

Table 4. Spearman and Pearson correlation coefficients among the  $C$ ,  $h$ ,  $T_1$ ,  $T_2$ , and  $ST$  of the top 30 highly cited patents.

	Spearman					
		$T_1$	$T_2$	$ST$	$C/P$	$h$
Pearson						
$T_1$		1	0.830*	0.988*	0.963*	0.892*
$T_2$		0.276	1	0.815*	0.839*	0.747*
$ST$		0.989*	0.363*	1	0.964*	0.905*
$C/P$		0.962*	0.410*	0.975*	1	0.959*
$h$		0.976*	0.347*	0.975*	0.985*	1

Note. \* Significant correlation at 0.01 level.

At the individual level, the differences in the average citation values and in the  $h$ -index values of the top 30 most cited papers were small. For the top 30 most cited patents, we determined that they could be divided to two groups: P1 to P7 had higher values of  $ST$  and  $T_1$  and a lower value of  $T_2$ , and P8 to P30 had approximately



the similar value of  $ST$ ,  $T_1$ , and  $T_2$ . The difference was due to the different citation types between the software and hardware patents.

Typically, an object receives trace metrics with a higher  $T_2$ , a higher  $ST$  that is lower than  $T_2$ , and a lower  $T_1$ . However, we observed that  $T_2$  was the lowest trace metric value of the top 30 most cited patents. A lower  $T_2$ , which considered the square of the citations in the tail part of the most cited patents, indicated that such citations were not comparable to those of their paper counterparts. However, although the average citation value in the tail part was lower than the  $h$ -index value, the values of the total citations in the tail part were usually higher than those of the total citations in the  $h$ -region (the core and the accessed parts). The rank-citation curves of the most cited papers, universities, and companies were gradual and had thick and long tails, but the corresponding rank-citation curves of patents were steep, with most citations accumulated in the  $h$ -area, and they had thin and short tails. Therefore, a lower  $T_2$  value represented a steep rank-citation curve, which is acceptable in patentometrics but is a symbol of irrelevance in bibliometrics.

We also determined that, at the individual level, all indicators showed significant correlations in both bibliometrics and patentometrics, demonstrating that all indicators, including traces, were effective indices for evaluation and cross-referencing.

## 4 Discussion and Conclusion

When the performance matrix proposed by Ye and Leydesdorff (2014) is extended to a primary matrix, secondary matrix, and submatrix, the traces of the three performance matrices  $T_1$ ,  $T_2$ , and  $ST$  can be applied to both bibliometrics and patentometrics. These trace metrics provide an integrated view of how citations are distributed by providing a scalar number. The performance in group level (i.e. a university or a company) or in individual level (i.e. a paper or a patent) can be evaluated by analyzing the value of the three traces.

Commonly used bibliometric indicators such as citation count and average citation are single point indicators, and they cannot accurately reflect variations in a rank-citation curve. Although the  $h$ -index includes publication and citation information simultaneously, it focuses on only the core region of the rank-citation curve. Trace metrics summarize the four parts of the rank-citation curve and thus provide a unique and integrated view. For example, the high number of low-citation papers resulted in Tsinghua University demonstrating a peak in Figure 2; however, P5 was special because it had few zero-citation papers (Figure 4). In our patentometric analyses (Figures 3 & 5), we determined that the different behaviors of trace metrics can be attributed to the different patent types (i.e. hardware or software patents).



Papers and patents in different fields might have different rank-citation curves but the same value of average citations and  $h$ -index. We observed that trace metrics could effectively distinguish between different patent types.

We observed that the differences in trace metrics were greater than those in the average citation values and in the  $h$ -index values (Figures 2–5). Because the trace metrics considered the square of information from different parts of the rank-citation curve, they were more sensitive to different types of publication and the citation status of various objects. In particular, in patentometrics, the values of patent citations are typically low, possibly resulting in commonly used indicators such as citation, average citation, and  $h$ -index not being sufficiently sensitive to indicate the difference.

For the trace metrics  $T_1$  and  $T_2$ , there was a negative term  $-\frac{P_z^2}{P}$ . We considered zero-citation publications and patents as a negative contribution to the total performance of an organization; this is because a proportion of research and development resources is consumed to conduct research projects and owning these zero-citation papers and patents, however, they do not have an impact on the related academia or market. If an organization has a large ratio of zero-citation papers or patents, which indicates that the usage of research and development resources is inefficient, it might receive a negative  $T_1$  or  $T_2$  value (e.g. Section 3.1 and Figure 3 show that IBM has a negative  $T_1$  value of approximately -5000). Therefore, these two indicators can facilitate decision makers in examining the impact efficiency of their organization.

If a university receives a negative  $T_1$  or  $T_2$  value, which means that it has produced few high-impact papers (but numerous irrelevant papers), we suggest that the governors of the university examine their research policy. Perhaps they should combine several less-impact projects into a more influential large project to advance their impact. Furthermore, if trace metrics are used to evaluate universities, the negative effect engendered by having many irrelevant papers can encourage universities to conduct substantial research or to publish comprehensive works, instead of several short and separated papers that increase the number of publications.

For patent owners, a negative trace metric value indicates imbalanced research and development distribution toward low-value patents. This might be tolerable for large enterprises because they might have sufficient capital to fabricate a long-term patent portfolio. However, for small businesses, it might indicate an impending financial failure to have such a negative value. By contrast, because the citing practice in patentometrics is different than in bibliometrics, and because certain patents receive low or zero citation despite being valuable, a negative value in  $T_1$  or  $T_2$  might be acceptable. We suggest that company managers regularly review their



own patents by using trace metrics. Because the maintenance fee for patents is a financial burden, managers can use trace metrics as a supplement to examine the value of their patents to determine which patents should be maintained.

The meaning of the negative term  $-\frac{P_z^2}{P}$  should be considered when using trace metrics. Trace metrics consider a zero-citation paper or patent as a negative appraisal. Therefore, before trace metrics are applied, clarifying how to evaluate a zero-citation paper or patent is advised.

A recent popular topic in bibliometrics and university evaluation is field normalization. This issue is usually discussed in university evaluation, and more and more global university ranking systems have adopted field normalization to reduce the field bias of publications and citations of different research-oriented universities. In our bibliometric test, we have already chosen a field so we could basically bypass this issue. Moreover, if we look into the subfields of computer sciences, we find that most of them have similar numbers of publications and citations therefore the field normalization issue could also be disregarded.

For our patentometric test, as we used the NBER categories, in which the smallest division is the computer software and hardware, to select our patent data, it is impossible for us to do field normalization in our patentometric analysis. However, for future research dealing with other fields, especially for fields that have significant bibliometric differences among their subfields, field normalization might be considered when evaluating the trace metric performance.

Our analysis reveals that trace metrics, which consider zero citation as a negative contribution, provide a unique view on the impact efficiency of an organization. We also determined that trace metrics exhibit different indicating behaviors between hardware patents and software patents, whereas commonly used indicators such as average citation and *h*-index have the same indication tendency between the different patent types. Because trace metrics are more sensitive and can provide the efficiency view, they are satisfactory substitutes for typical bibliometric and patentometric indicators, and they can help decision makers examine and adjust their policies.

## Acknowledgements

We acknowledge the National Natural Science Foundation of China (Grant No.: 71173187) and Jiangsu Key Laboratory Fund for financial support. We thank Ms. Danqi Shen, Ms. Helen F. Xue and Ms. Mei-Chun Lin for data collection and processing.





## Author Contributions

The work presented here was jointly conducted by F.Y. Ye (yye@nju.edu.cn), D.Z. Chen (dzchen@ntu.edu.tw), and M.H. Huang (mhhuang@ntu.edu.tw). F.Y. Ye developed the trace metrics method and wrote the methodology section. D.Z. Chen collected and analyzed the patent data and was responsible for the patent-related section. M.H. Huang collected and analyzed the bibliometric data and was in charge of the bibliometric section. All authors contributed to write the introduction section, discuss and interpret the results, as well as review the manuscript.

## References

- Anderson, T.R., Hankin, R.K.S., & Killworth, P.D. (2008). Beyond the Durfee square: Enhancing the *h*-index to score total publication output. *Scientometrics*, 76(3), 577–588.
- Bornheim, A., Bunn, J., Chen, J., Denis, G., Galvez, P., Gataullin, M., ... & Yuldashev, B.S. (2008). The CMS experiment at the CERN LHC. *Journal of Instrumentation*, 3(8), S08004.
- Bornmann, L., Mutz, R., & Daniel, H.D. (2010). The *h* index research output measurement: Two approaches to enhance its accuracy. *Journal of Informetrics*, 4(3), 407–414.
- Egghe, L. (2006). An improvement of the *h*-index: The *g*-index. *ISSI Newsletter*, 2(1), 8–9.
- García-Pérez, M.A. (2009). A multidimensional extension to Hirsch's *h*-index. *Scientometrics*, 81(3), 779–785.
- Garfield, E. (1972). Citation analysis as a tool in journal evaluation — Journals can be ranked by frequency and impact of citations for science policy studies. *Science*, 178, 471–479.
- Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences of the United States of America*, 102(46), 16569–16572.
- Huang, M.H., Chen, D.Z., Shen, D., Wang, M.S., & Ye, F.Y. (2015). Measuring technological performance of assignees using trace metrics in three fields. *Scientometrics*, 104(1), 61–86.
- Kuan, C.H., Huang, M.H., & Chen, D.Z. (2011). Positioning research and innovation performance using shape centroids of *h*-core and *h*-tail. *Journal of Informetrics*, 5(4), 515–528.
- Leydesdorff, L., & Bornmann, L. (2011). Integrated impact indicators compared with impact factors: An alternative research design with policy implications. *Journal of the American Society for Information Science and Technology*, 62(11), 2133–2146.
- Narin, F., Noma, E., & Perry, R. (1987). Patents as indicators of corporate technological strength. *Research Policy*, 16(2–4), 143–155.
- Schubert, A. (2009). Using the *h*-index for assessing single publications. *Scientometrics*, 78(3), 559–565.
- Waltman, L., & van Eck, N.J. (2012). The inconsistency of the *h*-index. *Journal of the American Society for Information Science and Technology*, 63(2), 406–415.
- Ye, F.Y., & Leydesdorff, L. (2014). The “academic trace” of the performance matrix: A mathematical synthesis of the *h*-index and the integrated impact indicator (I3). *Journal of the Association for Information Science and Technology*, 65(4), 742–750.
- Zheng, J., Zhao, Z., Zhang, X., Huang, M., & Chen, D. (2013). Influences of counting methods on country rankings: A perspective from patent analysis. *Scientometrics*, 98(3), 2087–2102.



## Appendix A

Table A1. List of top 30 highly cited paper in the field “computer sciences” from ESI.

No.	Authors	Journal	Publication year
P1	Robinson, M.D. et al.	<i>Bioinformatics</i>	2010
P2	Li, H. & Durbin, R.	<i>Bioinformatics</i>	2010
P3	Edgar, R.C.	<i>Bioinformatics</i>	2010
P4	Quinlan, A.R. & Hall, I.M.	<i>Bioinformatics</i>	2010
P5	Bullard, J.H. et al.	<i>BMC Bioinformatics</i>	2010
P6	Smoot, M.E. et al.	<i>Bioinformatics</i>	2011
P7	Waller, C.J. et al.	<i>Bioinformatics</i>	2010
P8	Wu, T.D. & Nacu, S.	<i>Bioinformatics</i>	2010
P9	Wang, L.K. et al.	<i>Bioinformatics</i>	2010
P10	Pruim, R.J. et al.	<i>Bioinformatics</i>	2010
P11	Quince, C. et al.	<i>BMC Bioinformatics</i>	2011
P12	Milne, I. et al.	<i>Bioinformatics</i>	2010
P13	Caporaso, J.G. et al.	<i>Bioinformatics</i>	2010
P14	Hadfield, J.D.	<i>Journal of Statistical Software</i>	2010
P15	Edgar, R.C. et al.	<i>Bioinformatics</i>	2011
P16	MacLean, B. et al.	<i>Bioinformatics</i>	2010
P17	Friedman, J. et al.	<i>Journal of Statistical Software</i>	2010
P18	McLaren, W. et al.	<i>Bioinformatics</i>	2010
P19	Hyatt, D. et al.	<i>BMC Bioinformatics</i>	2010
P20	Viechtbauer, W.	<i>Journal of Statistical Software</i>	2010
P21	Kembel, S.W. et al.	<i>Bioinformatics</i>	2010
P22	Danecek, P. et al.	<i>Bioinformatics</i>	2011
P23	Martin, D.P. et al.	<i>Bioinformatics</i>	2010
P24	Huang, Y. et al.	<i>Bioinformatics</i>	2010
P25	Pluskal, T. et al.	<i>BMC Bioinformatics</i>	2010
P26	Yu, N.Y. et al.	<i>Bioinformatics</i>	2010
P27	O’Boyle, N.M. et al.	<i>Journal of Cheminformatics</i>	2011
P28	Dweep, H. et al.	<i>Journal of Biomedical Informatics</i>	2011
P29	Baraniuk, R.G. et al.	<i>IEEE Transactions on Information Theory</i>	2010
P30	Robin, X. et al.	<i>BMC Bioinformatics</i>	2011

Table A2. Original records of top 30 cited patents in “computer software &amp; hardware”.

Patent	Patent number	Year	Assignee
P1	7665051	2010	Qimonda AG
P2	7802219	2010	Cadence Design Systems, Inc.
P3	7917877	2011	Cadence Design Systems, Inc.
P4	7712056	2010	Cadence Design Systems, Inc.
P5	7962867	2011	Cadence Design Systems, Inc.
P6	7992122	2011	GG Technology, Inc.
P7	7971160	2011	Fujitsu Semiconductor Limited
P8	7738971	2010	Ethicon Endo-Surgery, Inc.
P9	8306853	2012	Colts Laboratories
P10	7693720	2010	VoiceBox Technologies, Inc.
P11	7949529	2011	VoiceBox Technologies, Inc.
P12	8301709	2012	Google Inc.
P13	7685126	2010	Isilon Systems, Inc.



Patent	Patent number	Year	Assignee
P14	7716171	2010	EMC Corporation
P15	7650009	2010	Digimarc Corporation
P16	7809167	2010	Bell_Matthew
P17	8032409	2011	Accenture Global Services Limited
P18	7840537	2010	CommVault Systems, Inc.
P19	7827208	2010	Facebook, Inc.
P20	7643649	2010	Digimarc Corporation
P21	7647237	2010	MiniMed, Inc.
P22	7698160	2010	VirtualAgility, Inc
P23	7685254	2010	Pandya_Ashish A.
P24	7697719	2010	Digimarc Corporation
P25	7751596	2010	Digimarc Corporation
P26	7657849	2010	Apple Inc.
P27	7760905	2010	Digimarc Corporation
P28	7797204	2010	Balent_Bruce F.
P29	7653883	2010	Apple Inc.
P30	8200775	2012	Newsilike Media Group, Inc



This is an open access article licensed under the Creative Commons Attribution-NonCommercial-NoDerivs License (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

