A Two-Dimensional Approach to Performance Evaluation for a Large Number of Research Institutions

Chung-Huei Kuan
Department of Mechanical Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC. E-mail: d98522047@ntu.edu.tw

Mu-Hsuan Huang
Department of Library and Information Science, National Taiwan University, Taipei 10617, Taiwan, ROC. E-mail: mhhuang@ntu.edu.tw

Dar-Zen Chen
Department of Mechanical Engineering and Institute of Industrial Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC. E-mail: dzchen@ntu.edu.tw

We characterize the research performance of a large number of institutions in a two-dimensional coordinate system based on the shapes of their h-cores so that their relative performance can be conveniently observed and compared. The 2D distribution of these institutions is then utilized (1) to categorize the institutions into a number of qualitative groups revealing the nature of their performance, and (2) to determine the position of a specific institution among the set of institutions. The method is compared with some major h-type indices and tested with empirical data using clinical medicine as an illustrative case. The method is extensible to the research performance evaluation at other aggregation levels such as researchers, journals, departments, and nations.

Introduction

Performance evaluation of research institutions has long attracted interest, due to the practical needs of government officials, research funding agencies, institution administrators, and students. On the other hand, the h-index (Hirsch, 2005) has become a standard scientometric indicator in recent years for research performance evaluation, as is evident from the large number of related articles and its adoption by online databases such as Scopus and Web of Science. Quickly after its origination, the h-index has been extended and applied at various levels of aggregation such as departments, journals, institutions, or even nations.

Received May 16, 2011; revised October 12, 2011; accepted October 12, 2011

 • Published online 28 November 2011 in Wiley Online Library (wileyonlinelibrary. com). DOI: 10.1002/asi.21701

For the research performance evaluation of institutions, there are quite a number of papers disclosing various adaptations to the original h-index. Arencibia-Jorge, Barrios-Almaguer, Fernández-Hernández, and Carvajal-Espino (2008) applied the successive h-indices by Schubert (2007) to a hierarchy involving three levels of aggregation: researchers, departments, and institutions. From a similar yet different view point, Prathap (2006) proposed a two-level approach: a level-one h-index (h_1) which is the original h-index for the publications affiliated to the institution, and a level-two h-index (h_2) specifying that there are h_2 researchers in the institution, and each has an individual h-index at least h_2. Molinari and Molinari (2008a,b) decomposed the original h-index of an institution into the product of an impact index h_m and a factor related to the number of publications from the institution. Sypsa and Hatzakis (2009) further modified the impact index h_m by another factor so as to apply the modified h_m to comparing institutions of disparate sizes.

Although these papers all provide valuable insight into the comparison of institutional research performance, the adaptations mainly focus on the size or scale of the institutions, and what the adaptations incorporate or modify is still the original h-index, which is probably mostly criticized for being insensitive to the excessive citations of highly cited articles not accounted for by the h-index.

The foregoing adaptations' disregard of the h-index's insensitivity to excessive citations is probably due to the fact that it has already been covered by a large number of studies proposing various so-called h-type indices. Some examples of these h-type indices are the g-index (Egghe, 2006a,b), the h(2)-index (Kosmulski, 2006), the A-, R-, AR-indices

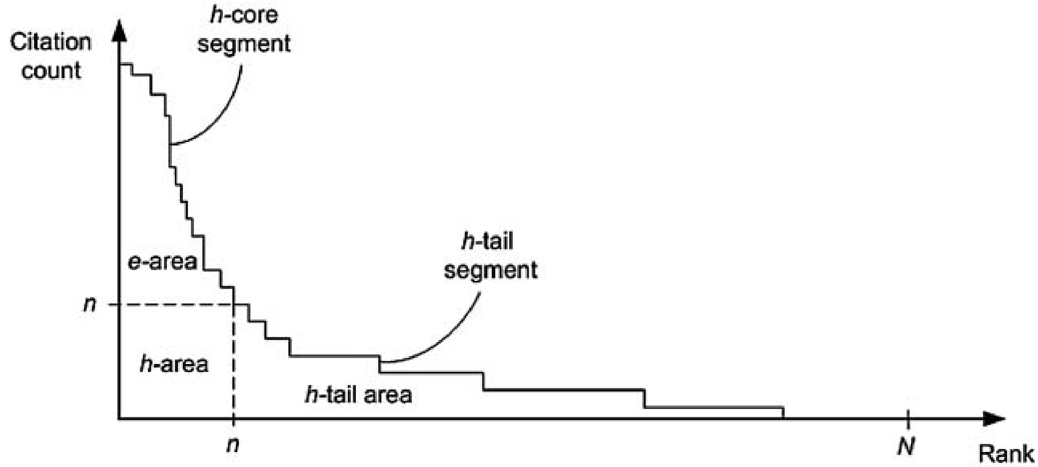

FIG. 1. Rank-citation curve of a fictitious institution's set of articles.

(Jin, 2007; Jin, Liang, Rousseau, & Egghe, 2007), the m-index (Bornmann, Mutz, & Daniel, 2008), the e-index (Zhang, 2009), the hg-index (Alonso, Cabrerizo, Herrera-Viedma, & Herrera, 2010), the q^2-index (Cabrerizo, Alonso, Herrera-Viedma, & Herrera, 2010), h-mixed synthetic indices S and T (Ye, 2010), and the w-index (Wu, 2010). Recent reviews and comparisons of these h-type indices can be found in Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2009), Egghe (2010), and Huang and Chi (2010).

We believe that, in institutional performance evaluation, whether the original h-index is supplemented by the number of high-level researchers or modified by the number of articles, the h-index's inherent insensitivity issue should not be ignored.

The above-mentioned h-type indices can be roughly categorized as those aiming to replace the original h-index (e.g., the g-, hg-, w-index), and those aiming to supplement the original h-index (e.g., the e-, A-, R-index). As the h-index has become a de facto standard and is readily available from on-line databases, we consider the latter as a more appealing approach (however, we are not implying that g-index and other indices are inferior).

For the h-type indices in the second category, they are almost without exception based on the sum of citation counts of the h-core (Rousseau, 2006), or the area size of the h-core under the curve manifesting the citation distribution of an individual's publications. This curve has accompanied the h-index since its origination but was given the name *rank-citation curve* only recently (Ye & Rousseau, 2010). More details about the rank-citation curve will be given below but at the moment, using a fictitious rank-citation curve shown in Figure 1, the e-index (Zhang, 2009) is exactly the size of the area marked as e-area, and the A-index and R-index (Jin, 2007; Jin et al., 2007) are equal to A_c/n and $\sqrt{A_c}$, respectively, where A_c is the combined size of the areas marked as e-area and h-area, Unfortunately, the sum of citation counts or area size is notorious in hiding details.

We believe that the shape of the rank-citation curve, rather than the area size underneath, is more appropriate in addressing the insensitivity issue of the h-index. The idea of using the geometry of the rank-citation curve was noticed earlier by Wohlin (2009) and recently by Kuan, Huang, and Chen (2011a). The study by Kuan et al. (2011a) focused on patent assignees, and the authors suggested that the rank-citation curve of an assignee with smaller h-index is located closer to the origin and therefore may run completely beneath the rank-citation curve of another assignee with greater h-index, implying that the former is outperformed by the latter. The authors then proposed two *shape descriptors* characterizing the segments of the rank-citation curve corresponding to an assignee's h-core and h-tail (Ye & Rousseau, 2010). The shape descriptors are then used to verify the geometric relationship among assignees' rank-citation curve segments, and thereby the assignees' relative performance with respect to their h-cores and h-tails.

Despite being suggested that the shape descriptors are extensible to research performance evaluation, the application of the shape descriptors to a large number of institutions is rather troublesome. The institutions have to be sorted first in accordance with their respective h-indices. Then the institutions' relative performance with respect to their h-cores and h-tails is further investigated by comparing their shape descriptors. In the process, a significant number of pairwise comparisons are required.

Additionally, and especially for a large number of institutions, an ideal method should allow us to gain an overall view of the relative performance for all institutions, instead of ranking them in a one-dimensional list where rank difference does not reveal much about the performance difference. More specifically, we envision the ideal method to be a 2D scheme where each institution's performance is represented by a characteristic point in a 2D coordinate system and, as such, a large number of institutions' performance can be simultaneously depicted and their relative performance then can be quickly determined by observing the relative positions of their characteristic points. In contrast to the monotonic linear ranking, the overall view offered by the 2D scheme allows us to see how institutions perform differently, where the difference lies, and

DOI: 10.1002/asi

how the institutions can be categorized, to name just a few colorful possibilities.

This paper describes our proposition for the characteristic point and the 2D scheme so that the insensitivity issue of the h-index is obviated while achieving the desired overall view for a large number of institutions. The proposition is then tested by empirical data. Even though this paper deals with the research performance evaluation for institutions, we believe that the method is equally applicable to research performance evaluation at different aggregation levels such as researchers, journals, departments, nations, etc.

Rank-Citation Curve and Shape Descriptor

Let an institution have a set of N affiliated articles $\{P_1, P_2, \ldots, P_{N-1}, P_N\}$ sorted in descending order of their respective citation counts $C(P_i)$, $1 \le i \le N$. The set of articles $\{P_1, P_2, \ldots, P_{N-1}, P_N\}$ is partitioned by the institution's h-index n into the set of more-cited n articles $\{P_1, P_2, \ldots, P_{n-1}, P_n\}$ and the set of less-cited and uncited $(N-n)$ articles $\{P_{n+1}, P_{n+2}, \ldots, P_{N-1}, P_N\}$, which are referred to as the institution's h-core (Rousseau, 2006) and h-tail (Ye & Rousseau, 2010), respectively.

The institution's rank-citation curve for the set of articles $\{P_1, P_2, \ldots, P_{N-1}, P_N\}$ is obtained by plotting and connecting the points $(i, C(P_i))$, $1 \le i \le N$, in a continuous or stepwise manner. A fictitious, stepwise rank-citation curve is depicted in Figure 1. The segments of the rank-citation curve over the h-core and the h-tail are referred to as the h-core and h-tail segments (Kuan et al., 2011a). The areas beneath the h-core and h-tail segments corresponding to citations received by the h-core and h-tail are referred to as the h-core area (Kuan et al., 2011a) and h-tail area (Ye & Rousseau, 2010). The h-core area (whose sizes is denoted as A_c) is further divided into the h-area (whose size is n^2) and the e-area (whose size is denoted as $A_e = A_c - n^2$) (Ye & Rousseau, 2010). The e-, h-, and h-tail areas are alternatively referred as h^2 upper, h^2 center, and h^2 lower by Bornmann, Mutz, and Daniel (2010).

In this paper we focus on the h-core area and the h-core segment only, ignoring the less-cited articles (e.g., those in the h-tail) is often considered an advantage of the h-index, and the more-cited articles (e.g., those in the h-core) should contribute more to an impact measure than less-cited articles (Egghe, 2010). This is also why the numerous h-type indices are proposed.

Considering that the rank-citation curve manifests the distribution of citations, Kuan et al. (2011a) proposed two shape descriptors, namely, the c- and t-descriptors, characterizing the shape of the h-core and h-tail segments, respectively. Since only the h-core is of concern here, we only provide the equation for the c-descriptor as follows:

$$c\text{-descriptor} = \sum_{i=1}^{n} C(P_i)\left(\frac{C(P_i)}{A_c}\right) = \frac{\sum_{i=1}^{n} C(P_i)^2}{\sum_{i=1}^{n} C(P_i)}, \quad (1)$$

and it should be easy to see that the c-descriptor satisfies the following inequality:

$$C(P_1) \ge c\text{-descriptor} \ge C(P_n) = n. \quad (2)$$

The c-descriptor is a weighted average of the heights of the points along the h-core segment or, equivalently, the weighted average of the citation counts of the h-core, where more weight is given to articles having more citations. As such, the c-descriptor has obviated the shortcoming of h-index as being insensitive to the excessive citations of highly cited articles. Additionally, according to Kuan et al. (2011a), given two individuals with sufficiently different h-indices, their c-descriptors allow us to verify that the one with greater h-index indeed outperforms the other with smaller h-index with respect to their h-cores. As to individuals with close or identical h-indices, their c-descriptors allow us to further differentiate their relative performance with respect to their h-cores.

The shape of the h-core segment is one aspect of the shape of the h-core area. From this perspective, then it would not be a surprise to note that the calculation of the c-descriptor as specified by Equation (1) is very similar to the way the shape centroid of the h-core area is obtained. As a matter of fact, we can derive the y-coordinate of the h-core centroid directly from the c-descriptor.

As illustrated in Figure 1, the h-core area can be considered as consisting of n rectangles, each having width 1, height $C(P_i)$ (therefore, area size $C(P_i)$), and having its centroid located at $(i - 0.5, C(P_i)/2)$, $1 \le i \le n$. According to geometry, the centroid of a planar shape divisible into a number of smaller shapes can be obtained as the weighted average of the centroids of these smaller constituent shapes. Therefore, the h-core centroid (c_x, c_y) can be obtained as follows:

$$c_y = \sum_{i=1}^{n} \frac{C(P_i)}{2}\left(\frac{C(P_i)}{A_c}\right) = \frac{1}{2}\frac{\sum_{i=1}^{n} C(P_i)^2}{\sum_{i=1}^{n} C(P_i)}; \quad (3)$$

$$c_x = \sum_{i=1}^{n} (i-0.5)\left(\frac{C(P_i)}{A_c}\right) = \frac{\sum_{i=1}^{n} (i-0.5)C(P_i)}{\sum_{i=1}^{n} C(P_i)}. \quad (4)$$

By comparing Equations (1) and (3), it should be easy to see the relationship between c_y, c-descriptor, and the h-index n is as follows:

$$C(P_1) \le c\text{-descriptor} = 2c_y \ge C(P_n) = n. \quad (5)$$

We therefore propose to use an institution's h-index (n) and c-descriptor ($2c_y$) as the characteristic point's x- and y-coordinates, where the h-index is considered a characterization of the shape of the h-core along the x direction and the c-descriptor, being exactly twice as large as c_y as indicated by Equation (5), is considered as a characterization of not

DOI: 10.1002/asi

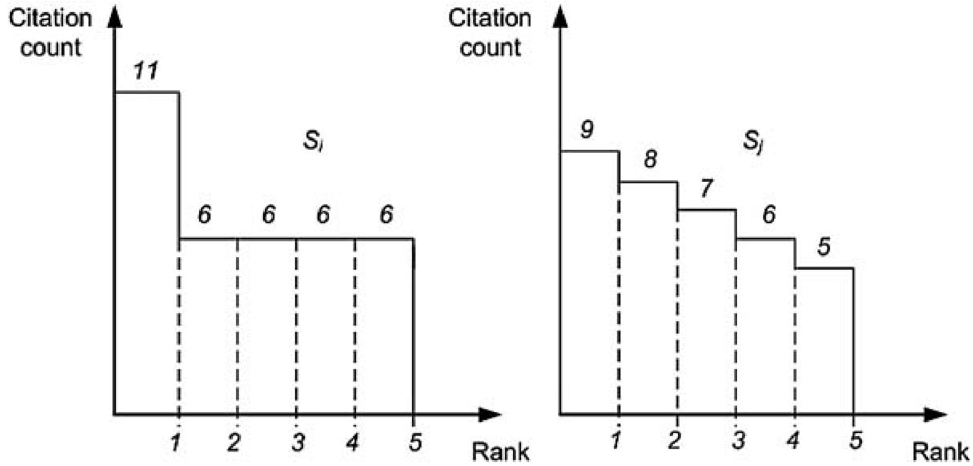

FIG. 2. h-Core areas of two fictitious institutions.

only the shape of the h-core segment but also the shape of the h-core area along the *y* direction.

It is easy to conjure up various candidates for the characteristic point, such as (n, A_e), (n, A_c), or other combinations involving the h-index and one of those h-type indices supplementing the h-index. However, we believe that the shape of the h-core segment or the shape of the h-core area is more discriminating than the area size. Figure 2 is a fictitious example showing the h-core areas of two institutions S_i and S_j. Their different *c*-descriptors, 7.57 for S_i and 7.28 for S_j, successfully indicate their h-core areas are differently shaped while their identical h-index 5, A_c 35, A_e (i.e., e-index) 10, A-index 7, and R-index 5.92 fail to provide differentiation.

However, one may ask, instead of $(n, 2c_y)$, why not directly use the h-core centroid (c_x, c_y) or (n, c_y) as the characteristic point. The latter is basically identical to $(n, 2c_y)$, and we adopt *c*-descriptor, $2c_y$, because it has a more intuitive range as specified by Equation (2). As to the former, this combination has already been employed and tested with empirical patent assignee data by Kuan, Huang, and Chen (2011b). According to Kuan et al. (2011b), (c_x, c_y) is indeed a viable choice. However, for research performance evaluation at a point in time such as this paper is intended to achieve, adopting c_x actually incurs a disadvantage.

According to Equations (1) and (3), if the shape of the h-core area is more skewed (e.g., S_i's h-core area in Figure 2) rather than extends gradually to the right (e.g., S_j's h-core area in Figure 2), c_y usually would be higher and c_x would be smaller, as more weight is given to articles of more citations. However, as pointed out by Kuan et al. (2011b), if the shape of an individual's h-core area is skewed to a certain degree, its c_x could be smaller than those having smaller h-indices. This is exactly where the confusion arises, as it would be difficult to tell, when one institution S_i's c_x is smaller than that of another institution S_j, whether S_i has a greater skewed h-core area or simply S_i has a smaller h-index.

In contrast, using the h-index *n* avoids such confusion entirely. Yet c_x does have its own merit, as suggested by Kuan et al. (2011b), for observing performance evolution over time. When one institution is found to have its c_x decrease or increase over a period of time, we can infer that, for the newly received citations within the time window, whether more is falling on the higher-ranking or lower-ranking articles.

We would like to point out that an institution's *c*-descriptor, and both c_x and c_y as specified by Equations (1), (3), and (4) can all be conveniently obtained along with the h-index within a single and same iteration through the institution's sorted set of articles.

The proposition of using an institution's h-index (*n*) and *c*-descriptor ($2c_y$) as the coordinates of its characteristic point, and plotting the characteristic points of multiple institutions in a 2D coordinate system can be further interpreted as follows. The h-index is commonly perceived as combining productivity (quantity) and impact (quality) in a single indicator, but leaving out the h-tail area and the e-area makes this combination less perfect. However, if we look at an institution's h-core only, the *c*-descriptor accurately manifests the impact side, while the h-index precisely reflects the productivity side of the institution's h-core. Therefore, by examining the relative positions of their characteristic points, we can see how the h-cores of multiple institutions perform differently in terms of their productivity and impact.

Research Data

The empirical data for testing our proposition are the 300 worldwide institutions having the greatest number of publications in the field clinical medicine recorded between the years 2008 and 2009. The 300 institutions are identified as follows. First, four types of documents (articles, reviews, research notes, and proceedings papers)[1] under 40 subject

[1]The four types of documents (articles, reviews, research notes, and proceedings papers) are used by Essential Science Indicators (ESI) for institutions (please see http://sciencewatch.com/about/met/core-ins/) and, for simplicity's sake, they are jointly referred to as articles.

DOI: 10.1002/asi

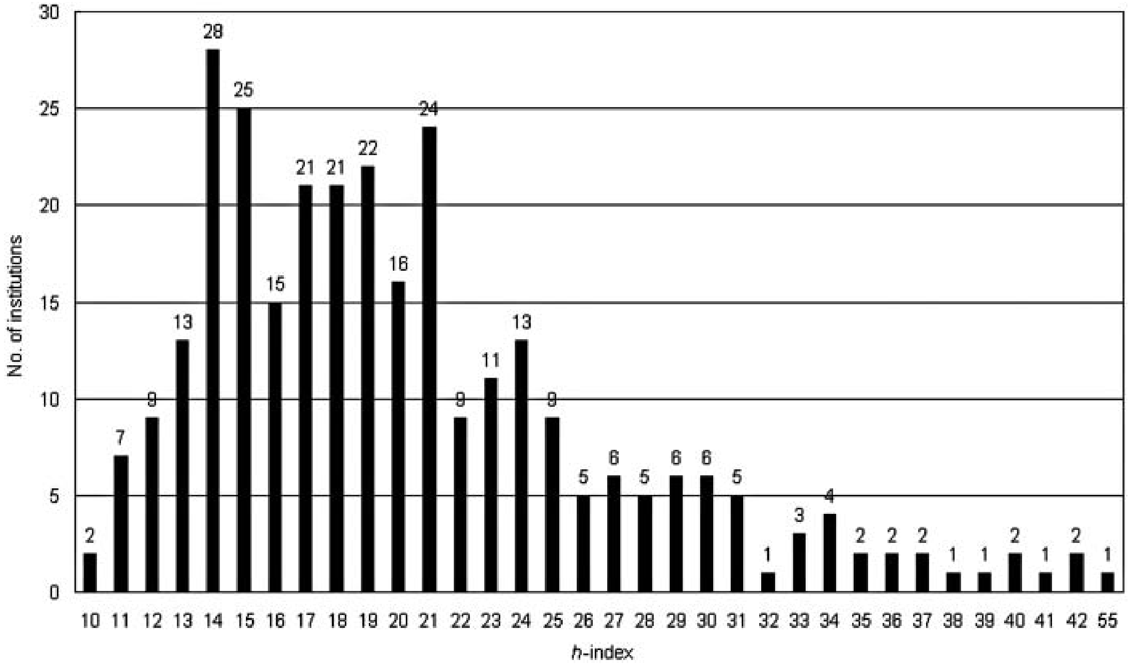

FIG. 3. The distribution of the 300 institutions among their h-indices.

categories[2] corresponding to the field "clinical medicine" whose index years (not publication years) are between 2008 and 2009 are collected from the SCI-EXPANDED and SSCI citation databases of Web of Science (updated on January 12, 2010). Then, the affiliated institutions of these documents are identified and the most productive 300 institutions are determined.

The field clinical medicine and the short time window are chosen on purpose. Even though some have suggested considering only articles at least 10 years old so that these articles have reached their citation potential (Molinari et al., 2008a), some have argued that a short time window is all right as long as all institutions are treated equally and for a field whose citations quickly attain a peak and whose citation potential is therefore fulfilled earlier (Rousseau, Yang, & Yue, 2010). Following the latter's reasoning, we have specifically chosen an even smaller time window so that a large number of institutions would have identical h-indices and our proposition can be tested to see how well these institutions are differentiated. As such, we found that, during this short time window, there were 468,573 articles from the 300 institutions, receiving 1,215,387 citations in total, and the h-indices of these institutions have a limited range from 55 (Harvard University with 11,268 articles and 38,835 citations) to 10 (Weizmann Institute of Science with 197 articles and 578 citations, and Princeton University with 129 articles and 427 citations).

As illustrated in Figure 3, there are only 34 different h-indices from these institutions, and therefore most of these h-indices are shared by multiple institutions. For example, there are 28 institutions with h-index 14, 25 institutions with h-index 15, and 24 institutions with h-index 21. There are also h-indices (e.g., 43–54) with no associated institutions and these h-indices are omitted in Figure 3.

c-Descriptor vs. e-, A-, and R-Indices

We choose the 28 institutions with h-index 14 as an example to see how well they are differentiated by the *c*-descriptor, and by three major h-type indices supplementing the h-index: the e-, A-, and R-indices. The relevant data of the 28 institutions are summarized in Table 1, sorted in descending order of their A_c's.

As shown in Table 1, the 28 institutions all have different *c*-descriptors and e-, A-, and R-indices. However, the rankings of the 28 institutions by the e-, A-, or R-indices achieve the same order as the e-, A-, and R-indices are all calculated from the A_c's. In other words, it makes no difference whether the e-, A-, or R-index is used in differentiating institutions of identical h-index.

As to the ranking by the *c*-descriptor, except the first two, the fifth, and the last three institutions, the order of the other 22 institutions are totally different from the order by the e-, A-, or R-index. The *c*-descriptor therefore must have

[2]The 40 subject categories include Allergy; Andrology; Anesthesiology; Cardiac & Cardiovascular Systems; Clinical Neurology; Critical Care Medicine; Dentistry, Oral Surgery & Medicine; Dermatology; Emergency Medicine; Endocrinology & Metabolism; Gastroenterology & Hepatology; Geriatrics & Gerontology; Gerontology; Health Care Sciences & Services; Hematology; Imaging Science & Photographic Technology; Integrative & Complementary Medicine; Medical Ethics; Medical Informatics; Medical Laboratory Technology; Medicine, General & Internal; Medicine, Legal; Medicine, Research & Experimental; Nursing; Obstetrics & Gynecology; Oncology; Ophthalmology; Orthopedics; Otorhinolaryngology; Pediatrics; Peripheral Vascular Disease; Radiology, Nuclear Medicine & Medical Imaging; Rehabilitation; Respiratory System; Rheumatology; Surgery; Transplantation; Tropical Medicine; Urology & Nephrology; and Psychiatry.

DOI: 10.1002/asi

TABLE 1. Relevant data for the 28 institutions with h-index 14.

Institution	A_c	e-Index	A-Index	R-Index	Rank by e, A, R*	c-Desc.	Rank by c*
Friedrich Schiller University of Jena	1,272	1,076	90.86	35.67	1	240.17	1
Universidade Federal de São Paulo	837	641	59.79	28.93	2	180.51	2
Université de la Méditerranée Aix-Marseille II	613	417	43.79	24.76	3	142.63	4
Martin Luther University of Halle-Wittenberg	599	403	42.79	24.47	4	177.87	3
Peking University	506	310	36.14	22.49	5	82.85	5
University of Tennessee – Health Science Center at Memphis	444	248	31.71	21.07	6	61.75	7
Justus-Liebig University Giessen	440	244	31.43	20.98	7	42.59	10
Université Toulouse III: Paul Sabatier	**436**	**240**	**31.14**	**20.88**	**8**	**42.39**	**11**
Université Joseph Fourier	**425**	**229**	**30.36**	**20.62**	**9**	**79.10**	**6**
Linköping University	**408**	**212**	**29.14**	**20.20**	**10**	**36.97**	**12**
University of Adelaide	**404**	**208**	**28.86**	**20.10**	**11**	**58.62**	**8**
University of Newcastle, Australia	**390**	**194**	**27.86**	**19.75**	**12**	**31.35**	**16**
Drexel University	**381**	**185**	**27.21**	**19.52**	**13**	**58.49**	**9**
Saarland University	**367**	**171**	**26.21**	**19.16**	**14**	**31.23**	**17**
University of Regensburg	365	169	26.07	19.10	15	32.65	14
University of Exeter	364	168	26.00	19.08	16	33.97	13
Technion – Israel Institute of Technology	355	159	25.36	18.84	17	30.82	18
Autonomous University of Barcelona	345	149	24.64	18.57	18	28.36	22
Tulane University	341	145	24.36	18.47	19	26.30	24
Queen's University Belfast	340	144	24.29	18.44	20	32.04	15
Louisiana State University – Baton Rouge	339	143	24.21	18.41	21	29.64	19
University of Nantes	336	140	24.00	18.33	22	29.54	20
Keio University	333	137	23.79	18.25	23	25.98	25
Chang Gung University	323	127	23.07	17.97	24	27.56	23
Nagoya University	321	125	22.93	17.92	25	28.36	21
University Louis Pasteur (Strasbourg I)	312	116	22.29	17.66	26	25.90	26
Kyushu University	303	107	21.64	17.41	27	23.20	27
Università Cattolica del Sacro Cuore	293	97	20.93	17.12	28	22.04	28

Note. *The Spearman correlation coefficient between the two rankings is 0.948, significant at 0.01 level (two-tailed).

extracted some different information from the documents of these institutions and the question therefore is which order, the one by the *c*-descriptor or the one by the e-, A-, or R-index, more accurately reflect the relative performance of the 22 institutions.

In order to present the comparison in a more readable manner, we chose seven institutions (Université Toulouse III: Paul Sabatier; Université Joseph Fourier; Linköping University; University of Adelaide; University of Newcastle, Australia; Drexel University; Saarland University) which are highlighted in Table 1 and ranked sequentially from the 8th to 14th places by the e-, A-, or R-index, and whose corresponding order by the *c*-descriptor seems to be most irregular. The h-core segments of the seven institutions are plotted in Figure 4. In the legend of Figure 4, each institution is associated with two numbers: the first being the rank by the e-, A-, or R-index, and the second being the rank by the *c*-descriptor.

As illustrated in Figure 4, the seven institutions can be separated into two groups: those having more skewed h-core segments and represented by hollow markers (Université Joseph Fourier; University of Adelaide; Drexel University), and those with less skewed h-core segments and represented by solid markers (Université Toulouse III: Paul Sabatier; Linköping University; University of Newcastle, Australia; Saarland University).

For the three institutions in the more skewed group, Drexel University is outperformed by University of Adelaide, as the former's h-core segment or area is almost completely inside or dominated (Kuan et al., 2011a) by that of the latter. On the other hand, even though a substantial part of the h-core segments or areas of University of Adelaide and Drexel University are slightly above that of Université Joseph Fourier, its first article has such a high visibility that Université Joseph Fourier can be considered to have the best relative performance. Such a scenario is accurately reflected in the rankings by the *c*-descriptor and by the e-, A-, or R-index.

For the four institutions in the less skewed group, it is easy to see that Université Toulouse III: Paul Sabatier outperforms Linköping University, which in turn outperforms Saarland University. As to University of Newcastle, Australia, even though the first three articles in its h-core receive fewer citations than the other three institutions, most of the rest of its h-core actually receives more citations. However, this is not enough to stop it lagging behind Université Toulouse III: Paul Sabatier and Linköping University, but is enough to make it surpass Saarland University. Again, such a scenario is accurately reflected in the rankings by the *c*-descriptor and by the e-, A-, or R-index.

In this example, if the two groups are compared separately, both the rankings by the e-, A-, or R-index and by the *c*-descriptor achieve the same relative order. However, if the institutions of the two groups are compared together, the two rankings produce very different results. For example, Université Toulouse III: Paul Sabatier is ranked at the first

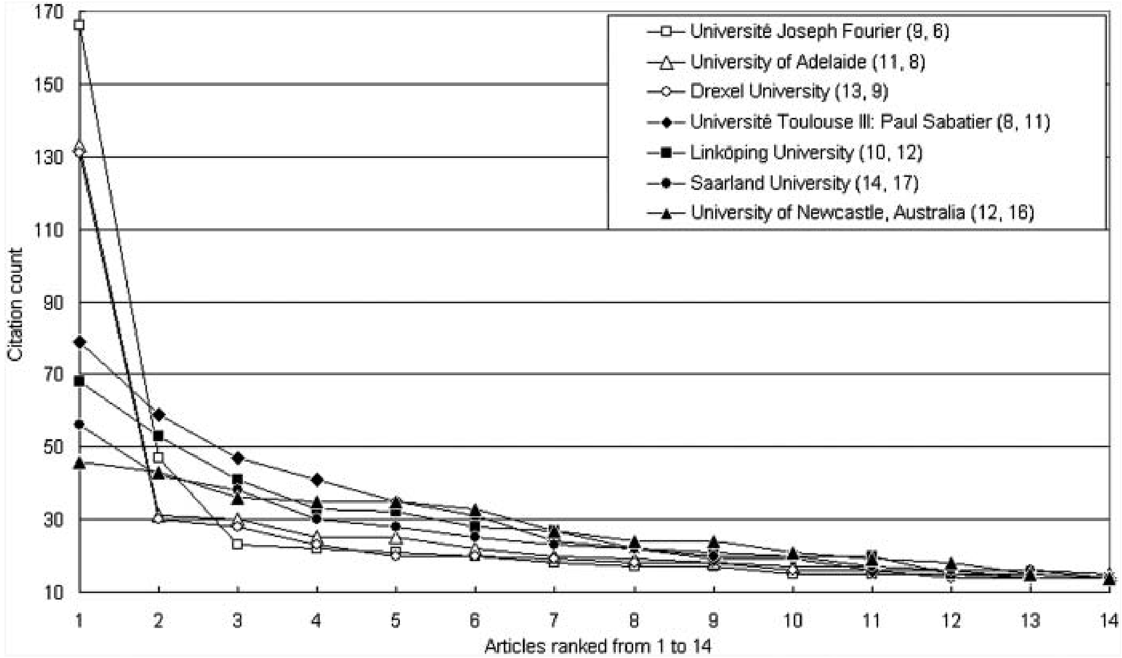

FIG. 4. h-Core segments for seven institutions with h-index 14.

place by the e-, A-, or R-index but, by the *c*-descriptor, it is only ranked at the 4th place.

We believe that the ranking by the *c*-descriptor actually reflects the relative performance of these institutions more accurately. This claim may seem implausible at first glance from Figure 4. However, the scales of *x*-axis and *y*-axis of Figure 4 are severely out of proportion. If Figure 4 is drawn to scale, its *y*-axis should be scaled up 20 times and the foregoing claim then should look more reasonable. With Figure 4 and Table 1 and together with Equation (1), we can see the reason why the institutions in the more skewed group are considered to have outperformed those institutions in the less skewed group. For the institutions in the more skewed group, their articles ranked at the 1st place receive 166, 133, and 131 citations, respectively, which account for at least 30% of their A_c's (ranging from 367 to 436). With such a significant proportion, the articles ranked at the 1st place from the more skewed group play a dominant role, causing the *c*-descriptors of the more skewed group to rise above those of the less skewed group. This scenario reveals a unique feature of the *c*-descriptor: a highly visible article is given more recognition than a number of mediocre articles. Due to this feature, for example, Université Toulouse III: Paul Sabatier is ranked only at the 4th place by the *c*-descriptor, but it actually has the greatest A_c.

This example may cause a false impression that, for institutions belonging to the same more skewed or less skewed group, the rankings by the e-, A-, or R-index, and by the *c*-descriptor would always achieve the same relative order. This is just coincidental. For a counterexample, as shown in Table 1, the University of Regensburg's A_c and *c*-descriptor are very close to those of Saarland University, and the skewness of its h-core segment must be close to that of Saarland University as well. However, their orders by the e-, A-, or R-index, and by the *c*-descriptor are reversed.

Despite that the *c*-descriptor excels the e-, A-, or R-index in ranking institutions with different degrees of skewness, the significant consistence in ranking institutions with similar degrees of skewness suggests that the two rankings are highly correlated. The respective Spearman correlation coefficients between the two rankings for the four most crowded sets of institutions with h-indices 14, 15, 21, and 19 are 0.948, 0.939, 0.964, and 0.966, all significant at the 0.01 level (two-tailed), indeed reflecting the high correlation.

Qualitative Categorization of Institutions

By using an institution's h-index (n) and *c*-descriptor ($2c_y$) as the *x*- and *y*-coordinates of the institution's characteristic point, the h-core performance for the 300 institutions can be simultaneously manifested by plotting their respective characteristic points in a 2D coordinate system as shown in Figure 5. Some of the characteristic points of the institutions listed in Table 1 are labeled with the names of the institutions.

In the Rank-Citation Curve and Shape Descriptor section, we have claimed that, for the 2D scheme as shown in Figure 5, the *x*-axis represents the productivity side while the *y*-axis represents the impact side of an institution's h-core. Then, as described in the *c*-Descriptor vs. e-, A-, and R-index section, for institutions with the same h-index (i.e., same productivity), those with greater *c*-descriptors (i.e., greater impact) are considered to have outperformed those with smaller *c*-descriptors. Correspondingly in Figure 5, the characteristic points of these institutions are scattered along a vertical line,

DOI: 10.1002/asi

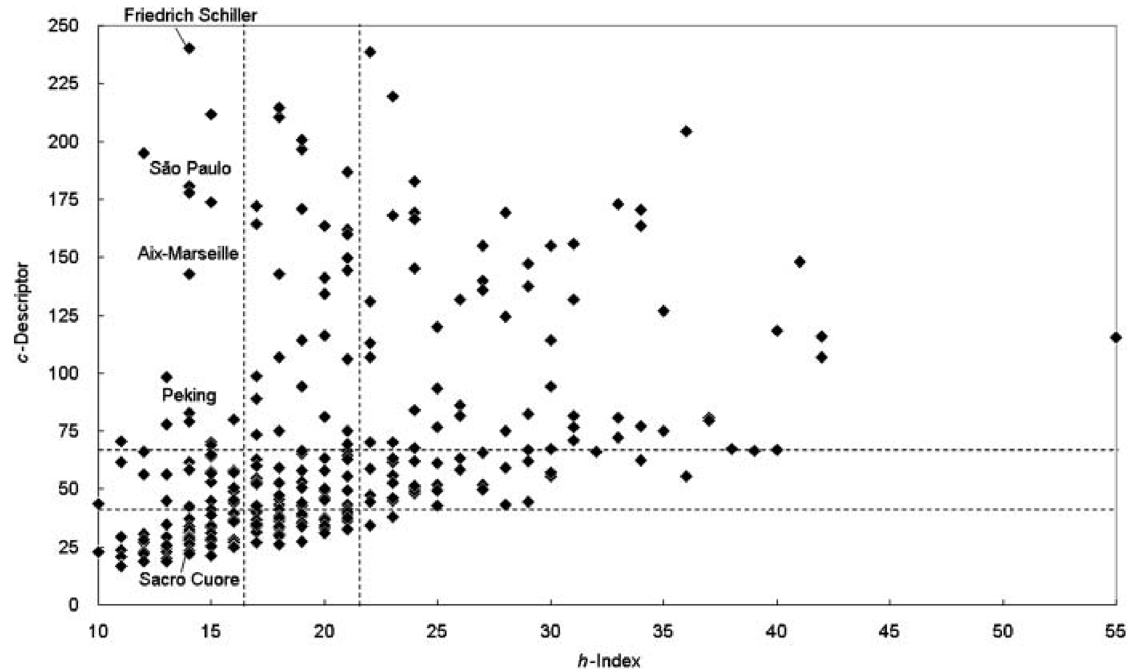

FIG. 5. The distribution of the 300 institutions' characteristic points.

with those higher above representing the institutions achieving greater performance. Similarly, for institutions with the same *c*-descriptor (i.e., same impact), those with greater h-indices (i.e., greater productivity) are considered to have outperformed those with smaller h-indices. Correspondingly in Figure 5, the characteristic points of these institutions are scattered along a horizontal line, with those to the far right representing the institutions achieving greater performance.

Following the foregoing reasoning, we can provide a quick qualitative categorization of the 300 institutions by overlaying an $l \times m$ grid on the 2D distribution of the characteristic points. The characteristic points are as such partitioned into $l \times m$ grid cells. For the institutions whose characteristic points are grouped in the same grid cell, they can be jointly considered as belonging to a same qualitative category, or as sharing a same performance nature.

Without losing generality, we choose $l = m = 3$ and the 300 institutions are separated into nine grid cells as follows. First, the h-indices of the 300 institutions are sorted, and a first h-threshold 16.5 and a second h-threshold 21.5 can be determined, where about 1/3 institutions (or 99 to be exact) having their h-indices <16.5, and about 1/3 institutions (or 97 to be exact) having their h-indices >21.5. Similarly, by sorting the *c*-descriptors of these institutions we can also determine a first *c*-threshold 42.8 and a second *c*-threshold 67.0 where about 1/3 institutions (or 100 to be exact) having their *c*-descriptors <42.8, and about 1/3 institutions (or 100 to be exact) having their *c*-descriptors >67.0.

In other words, the first and second h-thresholds together separate the characteristic points laterally into three groups having substantially the same number of points, and the first and second *c*-thresholds also vertically separate the characteristic points into three groups having substantially the same number of points, thereby achieving the nine-cell grid as shown in Figure 5 by the dashed lines.

TABLE 2. Qualitative categorization of the 300 institutions.

Low productivity High impact (15 institutions)	Medium productivity High impact (29 institutions)	High productivity High impact (56 institutions)
Low productivity Medium impact (25 institutions)	Medium productivity Medium impact (36 institutions)	High productivity Medium impact (39 institutions)
Low productivity Low impact (59 institutions)	Medium productivity Low impact (39 institutions)	High productivity Low impact (2 institutions)

In passing, we conducted a Kruskal–Wallis test to see, for the three groups of institutions divided by the h-thresholds, whether the medians of their constituent institutions' *c*-descriptors are statistically identical. The resulting *p*-value is less than 0.001 and therefore the hypothesis has to be rejected.

With the *x*- and *y*-axes indicating various degrees of productivity and impact, Table 2 shows nine performance natures attributed to the institutions within the nine grid cells, with each table cell (a qualitative category) corresponding to a grid cell at the corresponding location in Figure 5. In Table 2 the terms low, medium, and high stand for below average, average, and above average performance, and the numbers of institutions belonging to the qualitative categories are provided.

As can be seen from the above, the 300 institutions can be quickly categorized qualitatively, thereby gaining an immediate understanding of the nature of their performance. For example, we can easily identify the 56 institutions in the high-impact/high-productivity category and understand that

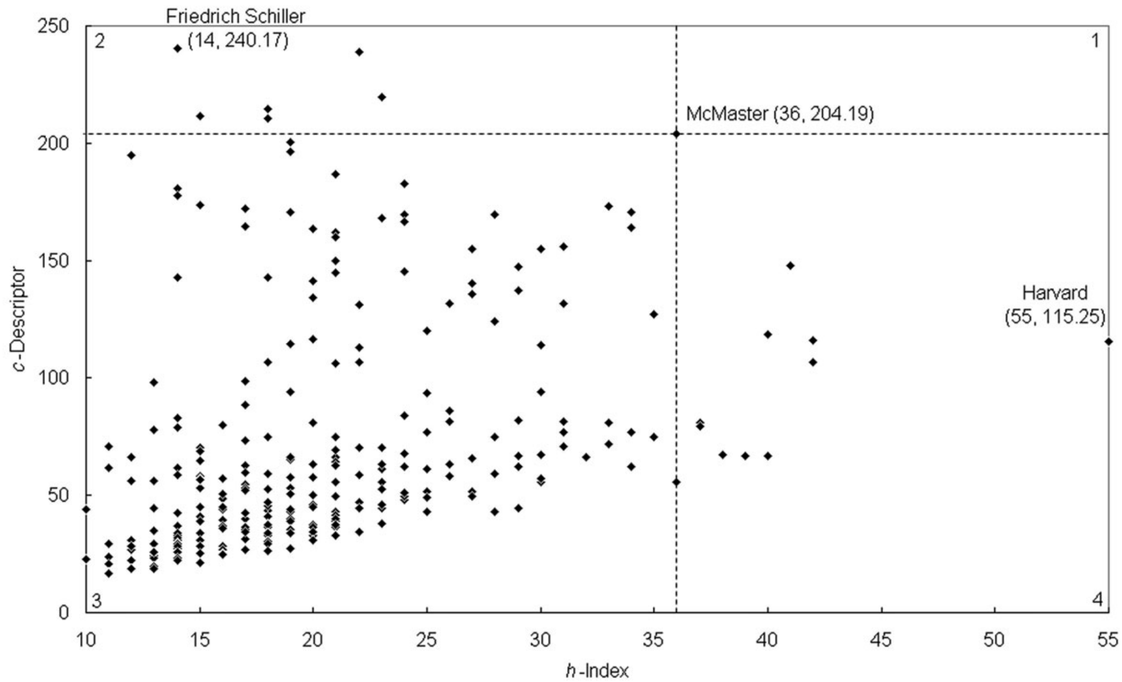

FIG. 6. Partitioning of the characteristic points into four quadrants relative to a reference point.

these 56 institutions have achieved the top one-third performance both in terms of their h-cores' impact and productivity. Similarly, we can also quickly see that there are only two institutions whose h-cores have achieved the top one-third performance in terms of productivity but only the last one-third performance in terms of impact. We therefore can infer that their h-core segments must be rather flat, compared to the other $95(=56+39)$ institutions also belonging to the high-productivity categories.

In this example the characteristic points are partitioned uniformly. But this is not required and an analyst can easily design various categorization schemes to suit their specific needs. For example, we can design specific *h*- and *c*-thresholds so as to determine institutions having achieve top 10% performance both in terms of their h-cores' productivity and impact.

Additionally, the above qualitative categorization should be valuable in monitoring an institution's performance evolution over a period of time. For example, if one institution is observed to have its characteristic point move from low-productivity/low-impact category, not diagonally to the medium-productivity/medium-impact category, but up to the low-productivity/medium-impact category, we can infer that most of its newly received citations fall on those already highly cited articles, thereby driving its characteristic point to move upward but providing little contribution to boost its h-index.

Determining the Performance of a Specific Institution

The qualitative categorization in the Qualitative Categorization of Institutions section provides some limited information on the relative performance of the 300 institutions. For the institutions whose characteristic points are aligned in a vertical or horizontal line (i.e., having identical degree of productivity or impact), we can quickly determine their relative performance. In addition, we can also infer that, for the institutions in the high-impact/high-productivity category must outperform those in the medium-impact/medium-productivity category which, in turn, outperform those in the low-impact/low productivity category.

Then, for institutions not fitting the above scenarios, how do we determine their relative performance or, more generally, how do we determine a specific institution's position among a group of institutions?

For simplicity's sake, we choose McMaster University, whose h-index is 36 and *c*-descriptor is 204.19, as an example. By treating its characteristic point as a reference point, the characteristic points of the other institutions are partitioned relative to the reference point into four quadrants numbered from 1 to 4 at the corners of Figure 6. For the characteristic points located on the left and lower segments of the partitioning lines, they are considered as belonging to the 3rd quadrant and, for those located on the right and upper segments of the partitioning lines, they are considered as belonging to the 1st quadrant.

As illustrated in Figure 6, there are 0, 6, 283, and 10 characteristic points in the 1st, 2nd, 3rd, and 4th quadrants, respectively. From McMaster University's point of view, for the institutions whose characteristic points are in its 3rd quadrant, McMaster University must have relatively superior performance, as it has both greater degrees of productivity and impact. Similarly, for the institutions whose characteristic points are in McMaster University's 1st quadrant, McMaster University must have relatively inferior

DOI: 10.1002/asi

TABLE 3. Relevant data for the 16 institutions whose characteristic points are in the 2nd and 4th quadrants of McMaster University.

Institution	Quadrant	h-Index	A_c	e-Index	A-Index	R-Index	c-Desc.
Harvard University	4	55	4,948	1,923	89.96	70.34	115.25
McMaster University	Reference	36	3,709	2,413	103.03	60.90	204.19
Johns Hopkins University	4	42	3,180	1,416	75.71	56.39	115.88
University of Washington – Seattle	4	41	3,128	1,447	76.29	55.93	147.88
University of Toronto	4	40	3,039	1,439	75.98	55.13	118.47
Mayo Clinic College of Medicine	4	42	2,995	1,231	71.31	54.73	106.73
University of Pennsylvania	4	37	2,473	1,104	66.84	49.73	80.74
University of California – Los Angeles	4	40	2,372	772	59.30	48.70	66.93
University of California – San Francisco	4	37	2,326	957	62.86	48.23	79.47
University of Texas – M. D. Anderson Cancer Center	4	39	2,308	787	59.18	48.04	66.55
Duke University	4	38	2,221	777	58.45	47.13	67.28
Wake Forest University	2	23	1,445	916	62.83	38.01	219.60
Friedrich Schiller University of Jena	2	14	1,272	1,076	90.86	35.67	240.17
Case Western Reserve University	2	22	1,244	760	56.55	35.27	238.65
University of Auckland	2	18	996	672	55.33	31.56	210.62
University of Leicester	2	18	935	611	51.94	30.58	214.57
Baylor University	2	15	699	474	46.60	26.44	211.77

performance, as it has both lower degrees of productivity and impact. For this example, we can then at least determine that McMaster University must be among the 17 institutions outperforming the other 283 institutions in terms of their h-core performance.

As to the 16 institutions whose characteristic points are in the 2nd and 4th quadrants, conclusions cannot be drawn confidently, as they have either greater productivity but lower impact, or greater impact but lower productivity, relative to McMaster University. The relevant data for the 16 institutions are summarized in Table 3, sorted in descending order of their A_c's. For comparison, the McMaster University's data are included as well.

According to Table 3, for example, Harvard University has the greatest h-index 55, yet its *c*-descriptor 115.25 is much smaller than McMaster University's 204.19. Therefore, we can expect that McMaster University must have a more skewed h-core segment so that its *c*-descriptor rises above that of Harvard University and so that Harvard University's characteristic point falls within McMaster University's 4th quadrant. Similarly, for Friedrich Schiller University of Jena, its *c*-descriptor 240.17 is greater yet its h-index 14 is much smaller than those of McMaster University. The characteristic point of Friedrich Schiller University of Jena is therefore located in the 2nd quadrant of McMaster University, as Friedrich Schiller University of Jena must have an even skewer h-core segment.

To verify the foregoing arguments, the respective h-core segments of the institutions falling within McMaster University's 2nd and 4th quadrants are plotted in Figures 7 and 8, respectively. In the figures, the h-indices of the institutions are connected to mark the boundaries of their h-cores and, for comparison, McMaster University's h-core segment is repeated in both figures. In order to be more readable, the *y*-axes of Figures 7 and 8 are log-scaled.

As shown in Figure 7, McMaster University outperforms the six institutions whose characteristic points are located in its 2nd quadrant, as these institutions' h-core segments or areas are completely beneath that of McMaster University. As to the 10 institutions whose characteristic points are located in McMaster University's 4th quadrant, we can tell from Figure 8 that McMaster University is only outperformed by Harvard University, as almost all of McMaster University's articles after the 3rd place receive significantly fewer citations than those of Harvard University, and the difference is too great to be compensated by the far greater citations received by McMaster University's first three articles. Other than Harvard University, McMaster University obviously outperforms the other nine institutions, as its first 22 articles all receive a greater number of citations and, after that, its articles receive fewer yet comparable number of citations.

The foregoing observation is consistent with the ranking by the institutions' A_c's (or R-indices) as shown in Table 3, which suggests that, when h-indices and *c*-descriptors cannot lead to a consistent conclusion to the relative performance among a group of institutions, their A_c's can step in and play a decisive role. The adoption of A_c in supplementing h-index and *c*-descriptor is not only proven to be accurate as shown in Figures 7 and 8, but also a reasonable choice. Intuitively, if two institutions' h-indices suggest that one is outperformed by the other while their *c*-descriptors suggest the other way around, the one with greater h-core citation count (i.e., A_c) should be considered to have superior performance.

Conclusion

We characterize an institution's h-core by using its h-index and *c*-descriptor as a characteristic point's *x*- and *y*-coordinates. The characteristic points of a large number of institutions then can be simultaneously depicted in a 2D

DOI: 10.1002/asi

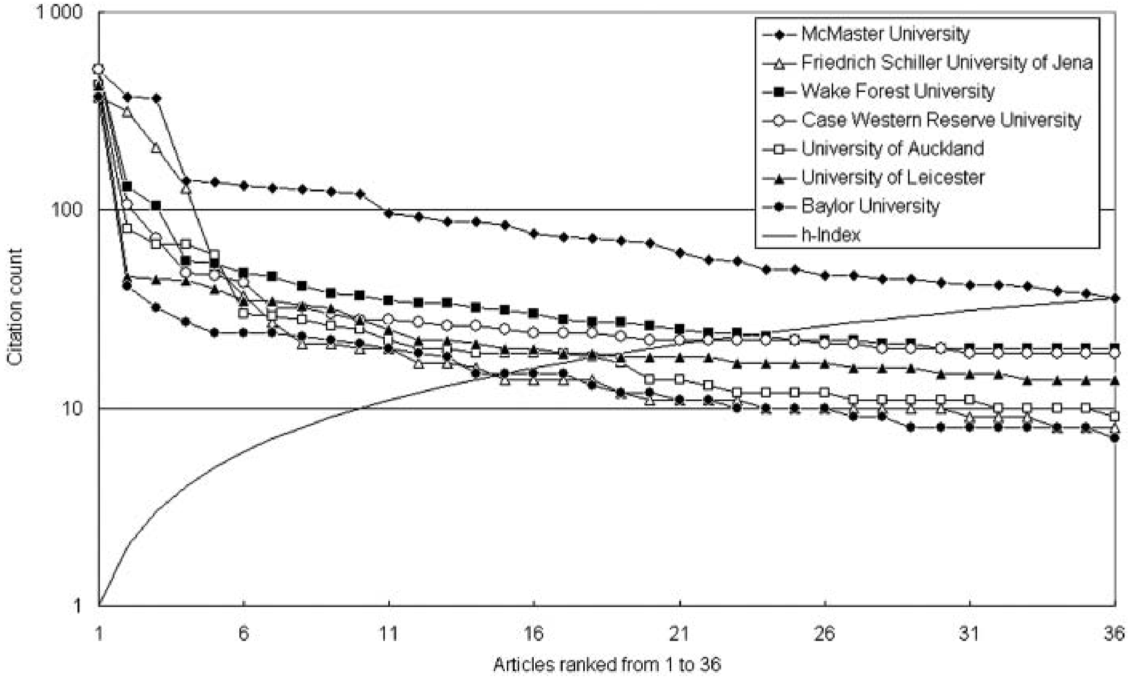

FIG. 7. h-Core segments of the institutions in McMaster University's 2nd quadrant (y-axis is log-scaled).

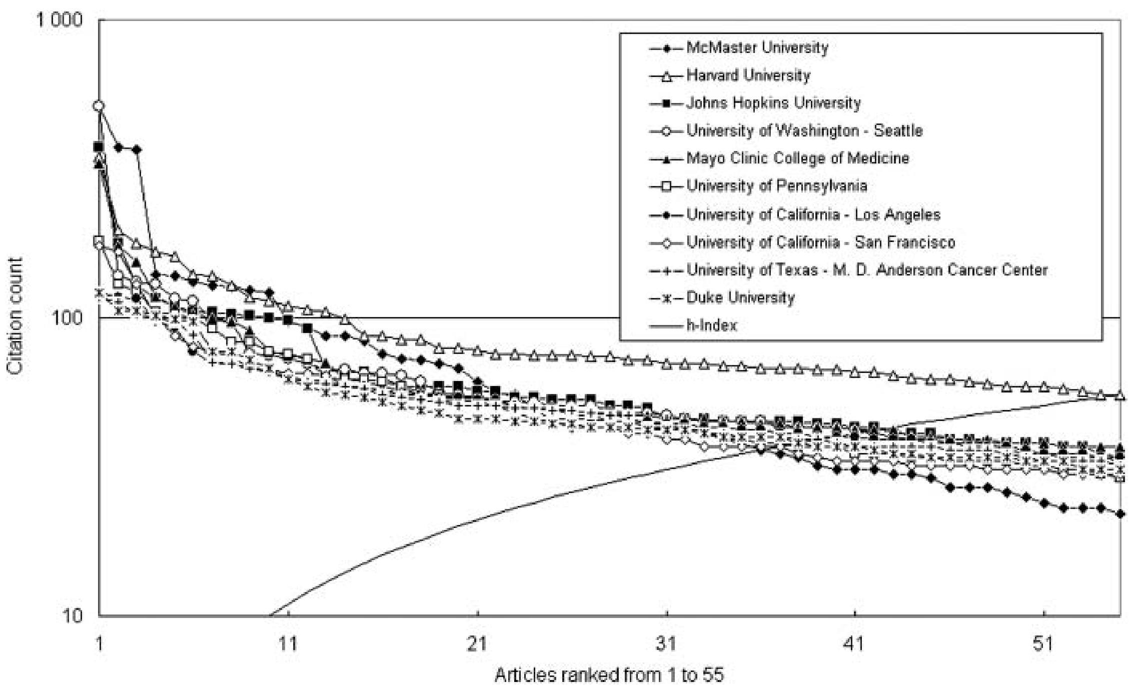

FIG. 8. h-Core segments of the institutions in McMaster University's 4th quadrant (y-axis is log-scaled).

coordinate system, thereby achieving an overall view of the institutions' h-core performance.

With the 2D distribution of characteristic points, we can quickly obtain a qualitative categorization of the institutions by partitioning the characteristic points into a number of grid cells with appropriately designed thresholds. For the institutions whose characteristic points are located in a same grid cell, they can be considered as sharing a same performance nature.

In addition, for institutions whose characteristic points are aligned vertically, horizontally, or diagonally, their relative performance can be quickly determined. As to institutions

DOI: 10.1002/asi

whose characteristic points are not vertically, horizontally, or diagonally aligned, their relative performance can be further determined by utilizing their A_c's (or R-indices).

Our method is proven by empirical data to be accurate and reliable. Even though it is found statistically to be highly correlated to the e-, A-, and R-indices, the *c*-descriptor adopted by our method successfully and accurately differentiates institutions of the same h-index but whose h-core segments have different degrees of skewness, while the e-, A-, and R-indices fail to do so.

While the paper focuses on institutional performance evaluation, the method, observations, and results are believed to be equally applicable to the research performance evaluation of individuals at various aggregation levels such as researchers, departments, journals, or even nations.

References

Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., & Herrera, F. (2009). *h*-Index: A review focused in its variants, computation and standardization for different scientific fields. Journal of Informetrics, 3, 273–289.

Alonso, S., Cabrerizo, F.J., Herrera-Viedma, E., & Herrera, F. (2010). *hg*-Index: A new index to characterize the scientific output of researchers based on the *h*- and *g*-indices. Scientometrics, 82, 391–400.

Arencibia-Jorge, R., Barrios-Almaguer, I., Fernández-Hernández, S., & Carvajal-Espino, R. (2008). Applying successive H indices in the institutional evaluation: A case study. Journal of the American Society for Information Science and Technology, 59, 155–157.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2008). Are there better indices for evaluation purposes than the *h* index? A comparison of nine different variants of the *h* index using data from biomedicine. Journal of the American Society for Information Science and Technology, 59, 830–837.

Bornmann, L., Mutz, R., & Daniel, H.-D. (2010). The *h* index research output measurement: Two approaches to enhance its accuracy. Journal of Informetrics, 4, 407–414.

Cabrerizo, F.J., Alonso, S., Herrera-Viedma, E., & Herrera, F. (2010). q^2-index: Quantitative and qualitative evaluation based on the number and impact of papers in the Hirsch core. Journal of Informetrics, 4, 23–28.

Egghe, L. (2006a). An improvement of the *h*-index: The *g*-index. ISSI Newsletter, 2, 8–9.

Egghe, L. (2006b). Theory and practise of the *g*-index. Scientometrics, 69, 131–152.

Egghe, L. (2010). The Hirsch-index and related impact measures. Annual Review of Information Science and Technology, 44, 65–114.

Hirsch, J.E. (2005). An index to quantify an individual's scientific research output. Proceedings of the National Academy of Sciences of United States of America, 102, 16569–16572.

Huang, M.-H., & Chi, P.-S. (2010). A comparative analysis of the application of *h*-index, *g*-index, and *A*-index in institutional-level research evaluation. Journal of Library and Information Studies, 8, 1–10.

Jin, B. (2007). The *AR*-index: Complementing the *h*-index. ISSI Newsletter, 3, 6.

Jin, B., Liang, L., Rousseau, R., & Egghe, L. (2007). The *R*-and *AR*-indices: Complementing the *h*-index. Chinese Science Bulletin, 52, 855–863.

Kosmulski, M. (2006). A new Hirsch-type index saves time and works equally well as the original *h*-index. ISSI Newsletter, 2, 4–6.

Kuan, C.-H., Huang, M.-H., & Chen, D.-Z. (2011a). Ranking patent assignee performance by *h*-index and shape descriptors. Journal of Informetrics, 5, 303–312.

Kuan, C.-H., Huang, M.-H., & Chen, D.-Z. (2011b). Positioning research and innovation performance using shape centroids of *h*-core and *h*-tail. Journal of Informetrics, 5, 515–528.

Molinari, J.-F., & Molinari, A. (2008a). A new methodology for ranking scientific institutions. Scientometrics, 75, 163–174.

Molinari, J.-F., & Molinari, A. (2008b). Mathematical aspects of a new criterion for ranking scientific institutions based on the *h*-index. Scientometrics, 75, 339–356.

Prathap, G. (2006). Hirsch-type indices for ranking institutions' scientific research output. Current Science, 91, 1439.

Rousseau, R. (2006). New developments related to the Hirsch index. Science Focus, 1, 23-25. (in Chinese). An English translation is available on-line at http://eprints.rclis.org/6376/.

Rousseau, R., Yang, L., & Yue, T. (2010). A discussion of Prathap's h_2-index for institutional evaluation with an application in the field of HIV infection and therapy. Journal of Informetrics, 4, 175–184.

Schubert, A. (2007). Successive *h*-indices. Scientometrics, 70, 201–205.

Sypsa, V., & Hatzakis, A. (2009). Assessing the impact of biomedical research in academic institutions of disparate sizes. BMC Medial Research Methodology, 9:33.

Wohlin, C. (2009). A new index for the citation curve of researchers. Scientometrics, 81, 521–533.

Wu, Q. (2010). The *w*-index: A measure to assess scientific impact by focusing on widely cited papers. Journal of the American Society for Information Science and Technology, 61, 609–614.

Ye, F.Y. (2010). Two *h*-mixed synthetic indices for the assessment of research performance. Journal of Library and Information Studies, 8, 1–9.

Ye, F.Y., & Rousseau, R. (2010). Probing the *h*-core: An investigation of the tail-core ratio for rank distributions. Scientometrics, 84, 431–439.

Zhang, C.-T. (2009). The *e*-index, complementing the *h*-index for excess citations. PLoS ONE, 4, e5429.

DOI: 10.1002/asi