Journal of Informetrics 5 (2011) 515–528

Contents lists available at ScienceDirect

Journal of Informetrics

journal homepage: www.elsevier.com/locate/joi

Positioning research and innovation performance using shape centroids of *h*-core and *h*-tail

Chung-Huei Kuan[a], Mu-Hsuan Huang[b], Dar-Zen Chen[c,*]

[a] Department of Mechanical Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC
[b] Department of Library and Information Science, National Taiwan University, Taipei 10617, Taiwan, ROC
[c] Department of Mechanical Engineering and Institute of Industrial Engineering, National Taiwan University, Taipei 10617, Taiwan, ROC

ARTICLE INFO

Article history:
Received 1 February 2011
Received in revised form 7 April 2011
Accepted 7 April 2011

Keywords:
Rank–citation curve
h-Core
h-Tail
Shape centroid
Patentometrics
Bibliometrics

ABSTRACT

We propose a novel yet practical method capturing an individual's research or innovation performance by the shape centroids of the *h*-core and *h*-tail areas of its publications or patents. A large number of individuals' relative performance with respect to their *h*-cores and *h*-tails can be simultaneously *positioned* and conveniently observed in two-dimensional coordinate systems. Two approaches are further proposed to the utilization of the two-dimensional distribution of shape centroids. The first approach specifically determines, within a group of individuals, those outperforming or being outperformed by a target individual. The second approach provides a quick qualitative categorization of the individuals so that the nature of their performance is revealed. Using patent assignees as an illustrative case, the approaches are tested with empirical patent assignee data.

1. Introduction

A curve manifesting the citation distribution of an individual's publications has accompanied the *h*-index since its origination (Hirsch, 2005) but was given the name *rank–citation curve* only recently (Ye & Rousseau, 2010). After appearing in Hirsch's original article, the rank–citation curve has been adopted by quite a number of *h*-index related articles, such as van Eck and Waltman (2008), Zhang (2009), Alonso, Cabrerizo, Herrera-Viedma, and Herrera (2009), Bornmann and Daniel (2009), and Bornmann, Mutz, and Daniel (2010), for graphically illustrating the authors' various propositions. There is also a web site providing *h*-index as well as its corresponding rank–citation curve based on the data of Google Scholar (Thor & Bornmann, n.d.).

Various measures about an individual's research performance can be derived from its rank–citation curve. For example, the total citation count corresponds to the area beneath the rank–citation curve and the cited patent count (i.e., the number of patents having been cited at least once) is where the rank–citation curve intersects the horizontal axis. In addition to these conventional performance measures, most of the recently developed *h*-type indices, such as the *g*-index (Egghe, 2006a, 2006b), $h^{(2)}$-index (Kosmulski, 2006), *A*-, *R*-indices (Jin, 2007; Jin, Liang, Rousseau, & Egghe, 2007), *m*-index (Bornmann, Mutz, & Daniel, 2008), *e*-index (Zhang, 2009), q^2-index (Cabrerizo, Alonso, Herrera-Viedma, & Herrera, 2010), *h*-mixed synthetic indices *S* and *T* (Ye, 2010) and *w*-index (Wu, 2010), could all be interpreted geometrically or better understood with the rank–citation curve. Recent reviews and comparisons of these *h*-type indices could be found, for example, in Alonso et al. (2009), Egghe (2010b), and Huang and Chi (2010).

* Corresponding author at: Department of Mechanical Engineering and Institute of Industrial Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei 10617, Taiwan, ROC. Tel.: +886 2 23692178; fax: +886 2 23631755.
E-mail addresses: d98522047@ntu.edu.tw (C.-H. Kuan), mhhuang@ntu.edu.tw (M.-H. Huang), dzchen@ntu.edu.tw (D.-Z. Chen).

1751-1577/$ – see front matter.
doi:10.1016/j.joi.2011.04.003

The nearly universal applicability of the rank–citation curve is not coincidental, as the rank–citation curve actually presents a snap shot of an individual's research performance up to a point of time. We therefore believe that, instead of being utilized merely as a graphical illustration tool, the rank–citation curve itself should be the basis for developing new indicators.

A number of articles have shifted their focus outside the highly cited *h*-core and incorporated, explicitly or implicitly, the entire rank–citation curve or the entire area underneath the rank–citation curve for research performance evaluation. Anderson, Hankin, and Killworth (2008) summed up every citation (i.e., every point along and under the rank–citation curve) with a weighting scheme based on the Durfee square into a so-called tapered *h*-index. García-Pérez (2009) proposed a multi-dimensional vector of *h*-indices, where the entire area beneath the rank–citation curve is quantized by a series of squares and each square's width corresponds to a component of the multi-dimensional vector. Bornmann et al. (2010) proposed to quantify three areas beneath the rank–citation curve: the area of lowly cited papers ranked behind the *h*-index (h^2 lower), the square area whose width is the *h*-index (h^2 center), and the area of excessive citations above the h^2 center (h^2 upper). Ye and Rousseau (2010) studied the ratio of the citations received by lowly cited *h*-tail papers to those received by the highly cited *h*-core papers as a so-called tail-core ratio and proposed a related *k*-index. Egghe (2010a) adopted a concept referred to as characteristic scores and scales (CSS) (Glänzel & Schubert, 1988) and proposed to summarize the research performance of a researcher by a number of points along its rank–citation curve, similar to the multi-dimensional vector of García-Pérez (2009).

The aforementioned approaches, except the tapered *h*-index, have the following shortcomings. Firstly, the proposed indicators are all based on the sizes of areas beneath the rank–citation curve but area sizes are notorious in hiding details. Secondly, for García-Pérez (2009)'s vector and Egghe (2010a)'s series, they are all directed to improve the accuracy of the original *h*-index, yet the indicators involved in their vector or series might be too numerous to maintain the simplicity advantage of the original *h*-index. In addition, even though the first few indicators in García-Pérez (2009)'s vector and Egghe (2010a)'s series do offer useful information but, for indicators derived from the far right of the rank–citation curve, there is little useful information. This is why García-Pérez (2009) proposed to truncate the indicators above a specific order from its multi-dimensional vector.

Following the same idea of adopting the entire rank–citation curve but focusing the application on patent assignees, Kuan, Huang, and Chen (2011) proposed to use the shape of the rank–citation curve for performance evaluation. The authors first 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.

From empirical data, Kuan et al. (2011) then found that, when the assignees' *h*-indices are sufficiently different, the above proposition would hold for most scenarios. To handle exceptions and the scenarios where the difference between *h*-indices is small, the authors proposed two shape descriptors, namely, the *c*-descriptor and *t*-descriptor, characterizing the segments of the rank–citation curve corresponding to an assignee's *h*-core and *h*-tail, respectively. The shape descriptors are then used to verify whether the geometric relationship among assignees' rank–citation curve segments, and thereby the assignees' relative performance with respect to their *h*-cores and *h*-tails, is indeed reflected by their respective *h*-indices.

Despite being proven by empirical data to be reliable, accurate, and robust, the application of the shape descriptors to a large number of assignees are quite cumbersome. The assignees have to be sorted first in accordance with their respective *h*-indices. Then, the assignees' relative performance with respect to their *h*-cores and *h*-tails is further investigated by comparing their *c*- and *t*-descriptors, respectively. In the process, a significant number of pair-wise comparisons are required and, most of all, it is difficult to gain an overall view of the relative positions of the assignees' performance.

We envision that an ideal methodology, applicable to individuals at various levels such as researchers, patent assignees, academic institutes, enterprises, or even nations, would be a two-dimensional scheme where an individual's research or innovation performance in terms of its publications or patents is represented by a characteristic point in a two-dimensional coordinate system, preferably with the characteristic point's horizontal and vertical coordinates capturing the productivity and impact sides of the individual's publications or patents, respectively. Through this two-dimensional scheme, a large number of individuals' research or innovation performance could be simultaneously depicted in the two-dimensional coordinate system and their relative performance could be immediately determined by observing the relative positions of their characteristic points.

This paper describes our proposition for such a characteristic point and, using patent assignees as an illustrative case, the application of our proposition to empirical patent assignee data.

2. Notations and shape centroids

Let $\{P_1, P_2, \ldots, P_{N-1}, P_N\}$ be an assignee's portfolio whose N patents are sorted in descending order of their respective citation counts $C(P_i)$, $1 \le i \le N$. The assignee's rank–citation curve is then obtained by plotting and connecting the points $(i, C(P_i))$, $1 \le i \le N$, together in a smooth or stepwise manner. An exemplary, stepwise rank–citation curve is depicted in Fig. 1.

The assignee's portfolio $\{P_1, P_2, \ldots, P_{N-1}, P_N\}$ is partitioned by its *h*-index n into the set of highly cited n patents $\{P_1, P_2, \ldots, P_{n-1}, P_n\}$ and the set of lowly cited and un-cited $(N-n)$ patents $\{P_{n+1}, P_{n+2}, \ldots, P_{N-1}, P_N\}$, which are referred to as the assignee's *h*-core (Rousseau, 2006) and *h*-tail (Ye & Rousseau, 2010), respectively. The *h*-tail is further divided into two subsets: the lowly cited patents $\{P_{n+1}, P_{n+2}, \ldots, P_{Nc-1}, P_{Nc}\}$ and the un-cited patents $\{P_{Nc+1}, P_{Nc+2}, \ldots, P_{N-1}, P_N\}$, where N_c

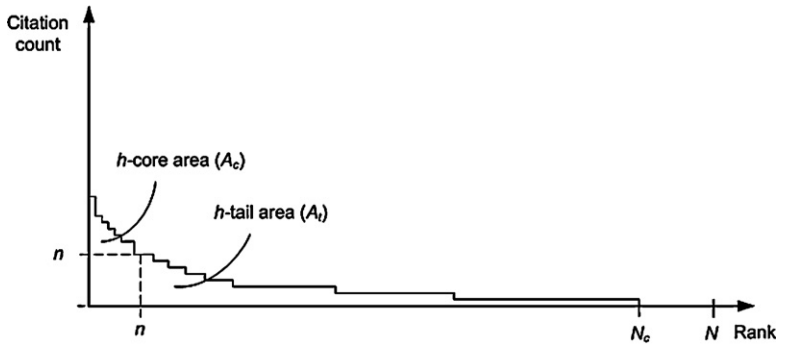

Fig. 1. Rank–citation curve of an assignee's portfolio.

is the number of cited patents. As un-cited patents are not associated with any impact information, we consider only cited patents and the term *h-tail* is referred only to the lowly cited patents $\{P_{n+1}, P_{n+2}, \ldots, P_{Nc-1}, P_{Nc}\}$ in this paper.

The areas beneath the rank–citation curve corresponding to citations received by the *h*-core and *h*-tail are referred to as *h*-core area (Kuan et al., 2011) and *h*-tail area (Ye & Rousseau, 2010), whose sizes are denoted as A_c and A_t, respectively. The *h*-core area 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). A_c is equal to the square of the *R*-index (Jin, 2007; Jin et al., 2007), and A_e is exactly the *e*-index (Zhang, 2009).

Considering the rank–citation curve manifests the distribution of citations of an assignee's portfolio, the *c*-descriptor and *t*-descriptor proposed by Kuan et al. (2011) are obtained as the weighted averages of the heights ($C(P_i)$) and the horizontal distances (i) of the points along the segments of the rank–citation curve over the *h*-core and the *h*-tail (hereinafter, *h*-core segment and *h*-tail segment), respectively:

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

$$t\text{-descriptor} = \sum_{i=n+1}^{N_c} i\left(\frac{C(P_i)}{A_t}\right) = \frac{\sum_{i=n+1}^{N_c} iC(P_i)}{\sum_{N_{i=n+1}}^{N_c} C(P_i)}. \tag{2}$$

According to Kuan et al. (2011), given two assignees with significantly different *h*-indices, the *c*- and *t*-descriptors allow us to verify that the one with greater *h*-index indeed outperforms the other with lesser *h*-index with respect to their *h*-cores and *h*-tails, respectively. As to assignees with close or identical *h*-indices, the *c*- and *t*-descriptors allow us to further differentiate their relative performance with respective to their *h*-cores and *h*-tails, respectively. The verification of the *c*- and *t*-descriptors' validity and their comparisons with conventional performance measures such as total citation count, number of cited patents, *h*-core citation count in ranking performance of patent assignees are covered in Kuan et al. (2011) with the same set of empirical data (see Section 3) as the current paper.

Interestingly, the calculations of the *c*- and *t*-descriptors as specified by Eqs. (1) and (2) are actually very similar to how the shape centroids of the *h*-core and *h*-tail areas are determined. As a matter of fact, given the *c*- and *t*-descriptors, we can immediately derive the *y*-coordinate of the *h*-core centroid and the *x*-coordinate of the *h*-tail centroid.

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

$$c_y = \sum_{i=1}^{n} \frac{C(P_i)}{2}\left(\frac{C(P_i)}{A_c}\right) = \frac{1}{2}\frac{\sum_{i=1}^{n} C(P_i)^2}{\sum_{i=1}^{n} C(P_i)} = \frac{1}{2}c\text{-descriptor}; \tag{3}$$

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

$$t_y = \frac{1}{2}\frac{\sum_{i=n+1}^{N_c} C(P_i)^2}{\sum_{i=n+1}^{N_c} C(P_i)}; \tag{5}$$

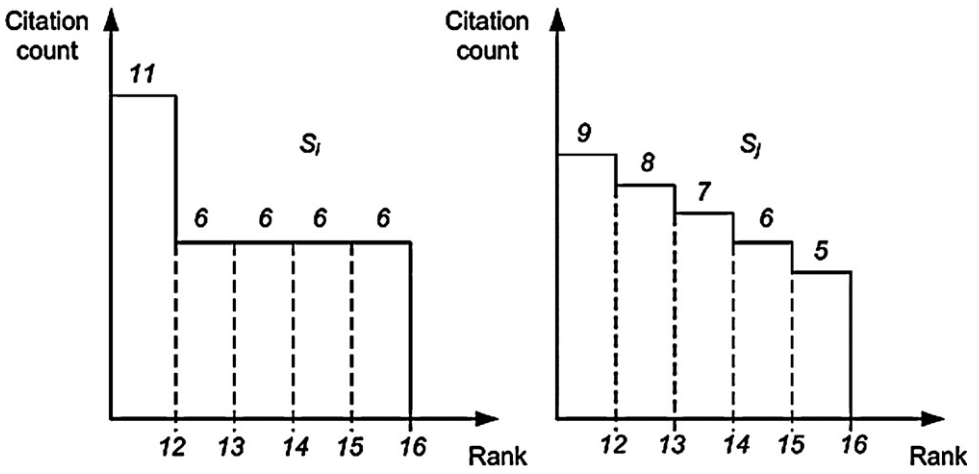

Fig. 2. *h*-Tail areas of two exemplary assignees.

$$t_x = \frac{\sum_{i=n+1}^{N_c}(i-0.5)C(P_i)}{\sum_{i=n+1}^{N_c}C(P_i)} = t\text{-descriptor} - 0.5. \tag{6}$$

As indicated by Eqs. (3)–(6), the *c*-descriptor is exactly twice as large as the *h*-core centroid's *y*-coordinate, c_y, and the *t*-descriptor is basically identical to the *h*-tail centroid's *x*-coordinate, t_x.[1] Additionally, the *x*-coordinate of the *h*-core centroid, c_x, could be considered as the *h*-core's *t*-descriptor, and the *y*-coordinate of the *h*-tail centroid, t_y, could be considered as the *h*-tail's *c*-descriptor.

The *x*- and *y*-coordinates of the *h*-core and *h*-tail centroids could be interpreted as follows. If the shapes of the *h*-core and *h*-tail areas are more skewed to the left, c_y and t_y would be higher, and c_x and t_x would be smaller, as more weight is given to the highly cited patents. In contrast, if the shapes of the *h*-core and *h*-tail areas extend smoothly to the right, c_x and t_x would be greater while c_y and t_y reflect a characteristic height for the slow slopes of the *h*-core and *h*-tail segments.

We propose to use the *h*-core and *h*-tail centroids as the characteristic points, as they are the geometric centers and thereby characterizations to the shapes of the *h*-core and *h*-tail areas, respectively. Even though it should not be hard to conjure up numerous other candidates for the characteristic points, such as (n, A_e), (n, A_c), (n, A_t), or various combinations of *h*-indices and area-based *h*-type indices, we believe that the shapes of the rank–citation curves or the shapes of the *h*-core and *h*-tail areas are more discriminating than the area sizes.

Secondly, despite the effectiveness of the *c*- and *t*-descriptors, unfortunately, there are scenarios where a single *c*- or *t*-descriptor could not achieve differentiation. Fig. 2 gives a simplified example where the *h*-tail areas of two assignees S_i and S_j, both with *h*-index 11, A_t 35, and N_c 16, are depicted. The *t*-descriptors for S_i and S_j are both 13.7 (therefore, t_x is 13.2), yet their different t_y's, 3.8 for S_i and 3.6 for S_j, successfully indicate their *h*-tail areas are differently shaped. In other words, when one coordinate of the shape centroids fails to provide discrimination, the other coordinate of the shape centroids could step in to fill the blank. Please note that the rounding position of the coordinates should be appropriately chosen depending on the data to be analyzed. In this example, coordinates are rounded to the first digit after the decimal point. For the rest of the paper, the coordinates are all rounded to integers for simplicity's sake, as integer coordinates are sufficiently discriminative.

Furthermore, the calculation of the *h*-core and *h*-tail centroids as specified by Eqs. (3)–(6) could be carried out with a single iteration of an assignee's sorted portfolio, and therefore could be obtained as a by-product when determining the *h*-index.

In addition, c_y and t_y are obviously related to the impact produced by an assignee's *h*-core and *h*-tail while c_x and t_x are citation-adjusted productivity measures for an assignee's *h*-core and *h*-tail. The *h*-core and *h*-tail centroids therefore fulfill our expectation for the characteristic point in capturing both the impact and productivity sides of an assignee's *h*-core and *h*-tail.

Finally, since more weight is given to highly cited patents, c_y has obviated the shortcoming of *h*-index as being insensitive to the highly cited patents.

Besides the advantages outlined above, we like to further point out that the usefulness of c_x and t_x is not limited to the static analysis conducted by this paper, and that they could actually provide valuable insight in the dynamic analysis of patent assignees. For example, if one assignee is observed to have its *h*-core centroid moving upward over a period of time. It would be difficult to tell whether the rising *y*-coordinate is resulted from a uniform increase of received citations from the entire *h*-core, or from a limited few highly cited patents' continuously receiving more citations. The two scenarios could

[1] If the sorted portfolio is arranged so that the rectangles are shifted to the left for a distance 0.5, the *t*-descriptor is identical to t_x.

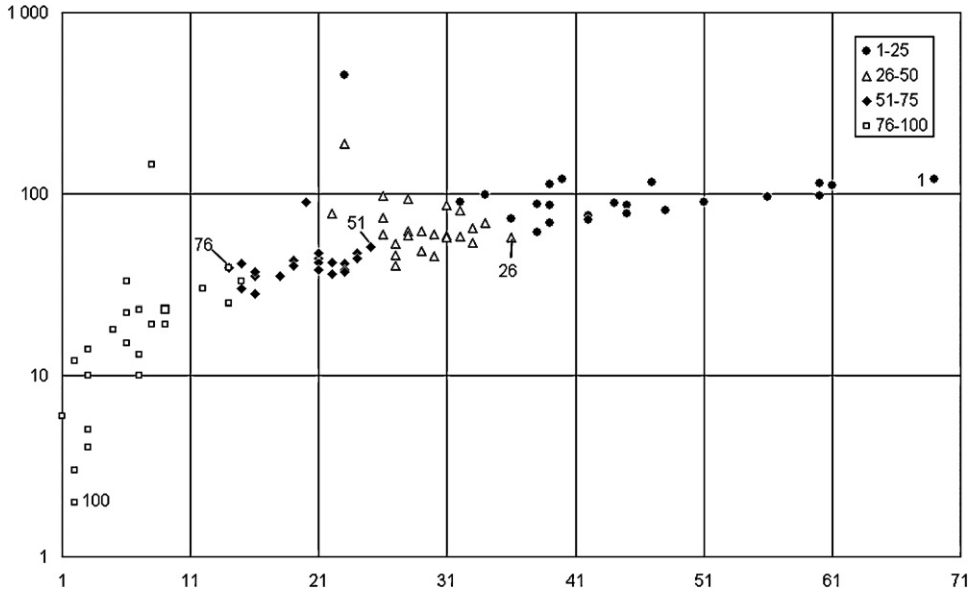

Fig. 3. Distribution of *h*-core centroids of the 100 assignees (*y*-axis is log-scaled).

be successfully differentiated by observing how c_x varies as, for the former, c_x would increase while, for the latter, c_x would decrease during the period of time.

3. Research data

The empirical data utilized by the paper is based on the 100 assignees having the greatest numbers of U.S. patents granted in the year 2009 (U.S. Patent and Trademark Office, 2009). These assignees' U.S. patents issued between 1976 and 2009 are then collected, and the respective *h*-indices are found to range from 161 (IBM with total 58,185 patents) to 3 (LG DISPLAY CO. LTD. with total 872 patents). From the diversity of these *h*-indices, the 100 assignees and their respective patent portfolios seem to constitute a representative set of data.

According to Kuan et al. (2011), 92 out of the 100 assignees have the ratios A_e/n^2 below 1.0, meaning that most assignees' *e*-areas are not greater than their respective *h*-areas (i.e., $A_e \le n^2$, or $A_c \le 2n^2$). For the 100 assignees, the ratios A_e/n^2 have a mean 0.58 with a standard deviation 0.37. On the other hand, the ratios A_t/n^2 have a mean 11.2 with a standard deviation 7, indicating that on the average most assignees have significantly long *h*-tail areas that are at least an order of magnitude greater than their respective *h*-areas. Therefore, an assignee's *h*-tail area constitutes such a huge portion of its productivity and impact to be ignored.

The *h*-core and *h*-tail centroids of the 100 assignees are obtained according to Eqs. (3)–(6) and plotted in Figs. 3 and 4, with log-scaled *y*-axis and log-scaled *x*-axis, respectively.

For comparison's sake, the centroids of the 100 assignees in Figs. 3 and 4 are plotted with different markers depending on their ranks by *h*-index.[2] For assignees ranked from the 1st to the 25th places, from the 26th to the 50th places, from the 51st to the 75th places, and from the 76th to the 100th places, four different markers, hollow squares, solid diamonds, hollow triangles, and solid circles are used, respectively. The centroids of the assignees ranked at the 1st (IBM), 26th (MITSUBISHI), 51st (NORTEL NETWORKS), 76th (HON HAI PRECISION), and 100th places (LG DISPLAY CO. LTD.) are also labeled with their respectively ranks.

At first glance, Figs. 3 and 4 may seem to be a bit confusing. It would be helpful to imagine that the *h*-cores and *h*-tails of the 100 assignees are combined and plotted altogether in the same diagrams and, then, their *h*-core and *h*-tail segments are removed, leaving the *h*-core and *h*-tail centroids in the diagrams. We will then have Figs. 3 and 4.

As illustrated in Fig. 3, for assignees with greater *h*-indices and therefore having their rank–citation curves located farther away from the origin (Kuan et al., 2011), their *h*-core centroids are usually positioned to the upper right of those assignees with smaller *h*-indices. However, there are a few exceptions where the *h*-core centroids for assignees with smaller *h*-indices are actually positioned either higher or more to the right than those assignees with greater *h*-indices. It is interesting to note that the *x*- and *y*-coordinates of the *h*-core centroids for the 100 assignees seem to follow a log-based trend line, whose R^2 is calculated to be 0.2314.

[2] For assignees of the same *h*-index, they are further sorted by their respective total citation counts.

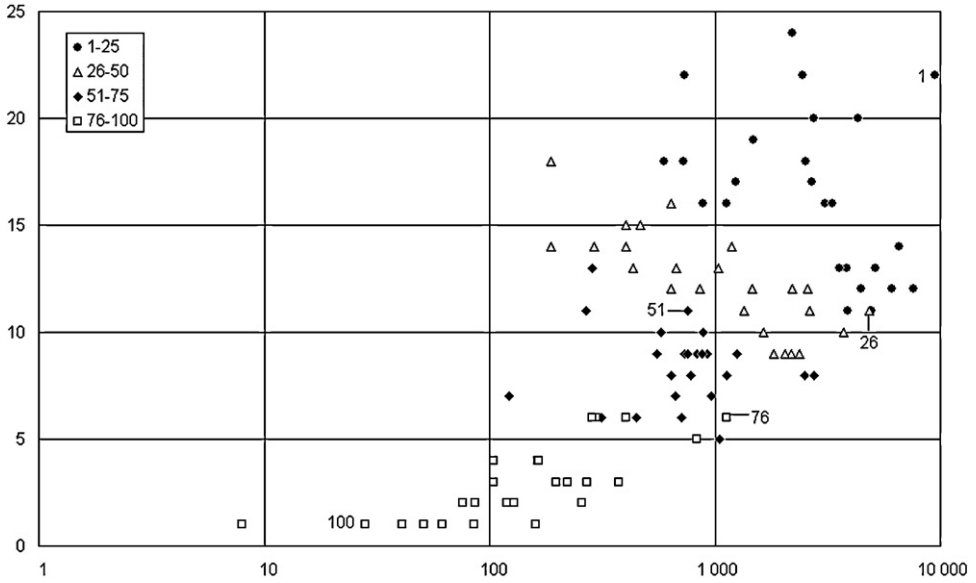

Fig. 4. Distribution of *h*-tail centroids of the 100 assignees (*x*-axis is log-scaled).

As illustrated in Fig. 4, from the positions of the *h*-tail centroids of assignees ranked at the 26th, 51st, and 76th places, the *h*-tail centroids are obviously more intermingled than the *h*-core centroids. However, we could still tell that, for assignees with greater *h*-indices, their *h*-tail centroids are usually positioned to the upper right of those assignees with smaller *h*-indices. The significantly more exceptions of the *h*-tail centroids arise from the significant variations of the assignees' *h*-tail areas. For the 100 assignees, the ratios A_t/n^2 could be as large as 38.1, and as small as 0.6 (Kuan et al., 2011). The *x*- and *y*-coordinates of the *h*-tail centroids for the 100 assignees also seem to follow an exponential-based trend line, whose R^2 is calculated to be 0.2003.

4. Positioning assignees by *h*-core centroids

4.1. First approach

For two assignees S_i and S_j, if S_i's *h*-core area is completely inside S_js, we refer to this scenario as S_i's *h*-core area being dominated by that of S_j, and we could claim that S_i is outperformed by S_j with respect to their respective *h*-cores, as S_i's *k*th patent always receives a less or equal number of citations to S_j's *k*th patent for all valid *k*'s, and the number of patents of S_i's *h*-core is also less than or equal to that of S_j.

If S_i's *h*-core area is dominated by S_j's, S_i's *h*-core centroid must be located to the lower left of S_j's *h*-core centroid. In other words, the *x*- and *y*-coordinates of S_i's *h*-core centroid must be both less than or equal to those of S_j's *h*-core centroid. The converse, even though not always true, provides us a hint to our first approach in utilizing the *h*-core centroids to position assignees' relative *h*-core performance.

In this approach, a target assignee's relative position among a group of assignees in term of their *h*-core performance is determined as follows. As an example and without losing generality, we choose the assignee whose *h*-index is ranked at the 51st place (NORTEL NETWORKS, with *h*-index 60) among the 100 assignees as our target. By treating the target's *h*-core centroid as a reference point, the *h*-core centroids of the rest of the assignees are partitioned relative to the reference point into four quadrants numbered from 1 to 4 at the corners of Fig. 5.

For the *h*-core centroids whose *x*- and *y*-coordinates both are greater than or equal to those of the reference point, they are considered as belonging to the 1st quadrant and represented by circle markers. For the *h*-core centroids whose *x*- and *y*-coordinates both are less than or equal to those of the reference point, they are considered as belonging to the 3rd quadrant and represented by square markers. As to *h*-core centroids having smaller *x*-coordinate but greater *y*-coordinate than those of the reference point, they are considered as belonging to the 2nd quadrant and represented by triangle markers. Similarly, *h*-core centroids having greater *x*-coordinate but smaller *y*-coordinate are considered as belonging to the 4th quadrant and represented by diamond markers. It is possible that an assignee has its *h*-core centroid located at the exactly same spot as the reference point and the *h*-core centroid therefore does not belong to any quadrant. This exception will be covered later.

Additionally, a hollow marker indicates that the *h*-core centroid's corresponding *h*-index is less than or equal to that of the target, and a solid marker indicates the opposite.

Therefore, as illustrated in Fig. 5, the *h*-core centroids in the 1st quadrant all have corresponding *h*-indices greater than the target's *h*-index, and those in the 3rd quadrant all have corresponding *h*-indices less than or equal to the target's *h*-index. As

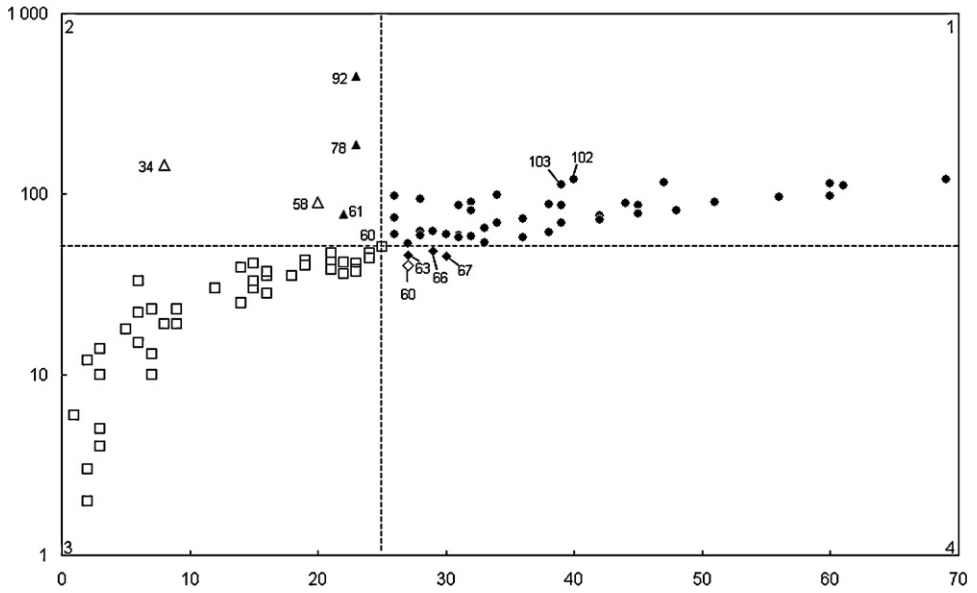

Fig. 5. Distribution of *h*-core centroids relative to a reference point (*y*-axis is log-scaled).

to those located in the 2nd and 4th quadrants, both conditions are present and they are labeled by their respective *h*-indices (not the ranks!). In other words, for assignees with greater *h*-indices than that of the target, none of their *h*-core centroids appears in the reference point's 3rd quadrant and, for assignees with *h*-indices less than or equal to that of the target, none of their *h*-core centroids appears in the reference point's 1st quadrant.

The idea of involving the *h*-index here is to use it as supporting evidence in inferring the domination relationship among the assignees' *h*-core areas. As mentioned in the beginning of this section, if S_i's *h*-core centroid is located to the lower left (i.e., 3rd quadrant) of S_j's *h*-core centroid, it is very possible, but not always true, that S_i's *h*-core area is dominated by S_j's. If additionally S_i's *h*-index is also less than or equal to S_j's *h*-index, the possibility that S_i's *h*-core area is dominated by that of S_j should be even higher.

As suggested by Fig. 5, it indeed seems that the assignees whose *h*-core centroids are located in the reference point's 1st or 3rd quadrant always have *h*-indices greater than or no greater than that of the target. However, for the 100 assignees, there is indeed one exception. MEDTRONIC INC. (ranked at the 12th place) has *h*-index 103, greater than SEMICONDUCTOR ENERGY LAB. (ranked at the 13th place)'s *h*-index 102, yet MEDTRONIC INC.'s *h*-core centroid (39, 112) is in the 3rd quadrant of SEMICONDUCTOR ENERGY LAB.'s *h*-core centroid (40, 121). The two assignees' *h*-core centroids are marked by their *h*-indices in Fig. 5 and their *h*-core areas are depicted in Fig. 6.

As illustrated in Fig. 6, SEMICONDUCTOR ENERGY LAB. indeed has more patents receiving more citations than MEDTRONIC INC., and their *h*-core centroids accurately indicate that SEMICONDUCTOR ENERGY LAB. outperforms MEDTRONIC INC. with respect to their *h*-cores while their *h*-indices mistakenly suggest otherwise.

With the confidence built from the foregoing observation, we propose that, if an assignee's *h*-core centroid is located in the reference point's 1st or 3rd quadrant, the assignee outperforms or is outperformed by the target with respect to their *h*-cores. The incorporation of the *h*-indices in the foregoing discussion could be omitted in real-life application for simplicity's sake. However, exceptions, if exists, are very possible to occur for those assignees whose *h*-core centroids are at close proximity to the target, as suggested by the example of MEDTRONIC INC. and SEMICONDUCTOR ENERGY LAB. Therefore, additional verification could be exercised by further analyzing the assignees whose *h*-core centroids adjacent to the reference point. Additionally, the assignee MONSANTO ranked at the 52nd place is calculated to have its *h*-core centroid (after rounding to integers) coinciding with the reference point. We could either claim that MONSANTO and the target, NORTEL NETWORKS, have comparable performance or conduct further analysis to see if they can be differentiated. According to Kuan et al. (2011) which is based on the same set of empirical data, if the difference between two *h*-indices is at least 8, they could be considered significantly different. This could be used as hint to decide what *h*-core centroids are considered to be at close proximity to the target.

What we need to deal with now is the assignees whose *h*-core centroids are located in the 2nd and 4th quadrants in Fig. 5. The relevant data of these assignees and the target is summarized in Table 1.

For PIONEER HI-BRED, it has a much smaller *h*-index 34 than that of the target, NORTEL NETWORKS. We therefore would expect that PIONEER HI-BRED's *h*-core centroid is located to the left of and beneath the reference point (i.e., NORTEL NETWORKS' *h*-core centroid). Yet, this is not what is shown in Fig. 5. According to Eq. (3), PIONEER HI-BRED must have a number of highly cited patents in its *h*-core, much higher than those of NORTEL NETWORKS' *h*-core. After examining their data, it is indeed the case. The first few patents in PIONEER HI-BRED's *h*-core receive 502, 496, 481, and 415 citations,

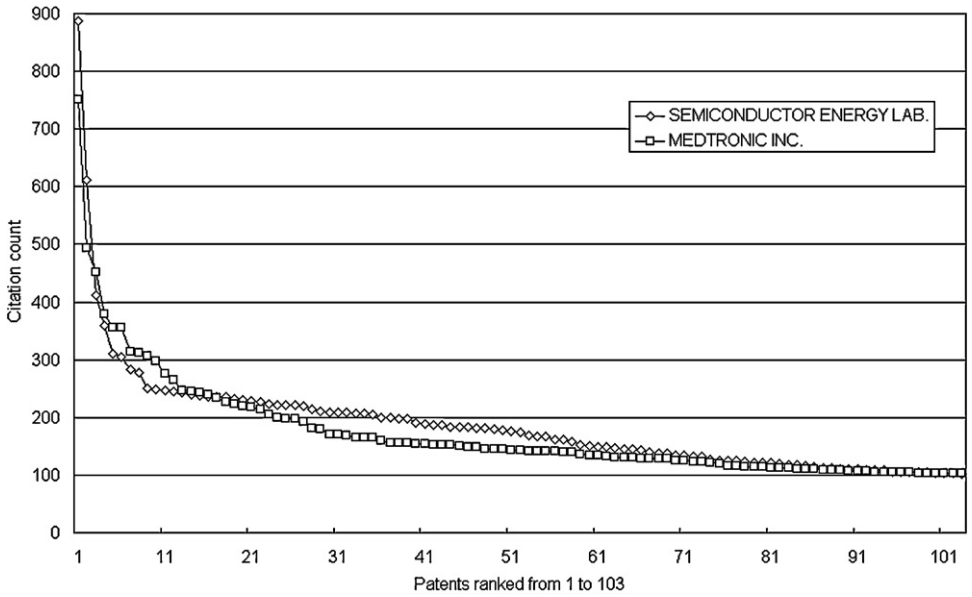

Fig. 6. *h*-Core areas of MEDTRONIC INC. and SEMICONDUCTOR ENERGY LAB.

respectively, while the highest ranking patent of NORTEL NETWORKS has received only 205 citations, and the subsequent patents of PIONEER HI-BRED all receive less citations than those of NORTEL NETWORKS' corresponding patents. In other words, the citation distribution of PIONEER HI-BRED skewed enough so as to "move" its *h*-core centroid above NORTEL NETWORKS' *h*-core centroid.

As to CANON INC., it has a much greater *h*-index 92 than that of NORTEL NETWORKS. We would expect that CANON INC.'s *h*-core centroid is located to the right of the reference point (i.e., NORTEL NETWORKS' *h*-core centroid). Again, this is not what is shown in Fig. 5. According to Eq. (3), the citation distribution of CANON INC.'s *h*-core must be skewed so severely that its *h*-core centroid is located to left of NORTEL NETWORKS' *h*-core centroid. This statement is also true after examining their respective data. The first few patents in CANON INC.'s *h*-core receive 1,950, 1,690, 1,649, 1,571, 1,549, 1,515, 1,472, and then abruptly down to 263 citations. QUALCOMM also exhibits similar skewness. The first two patents in QUALCOMM.'s *h*-core receive 1,139 and 974 citations, and the citations drop to 193 citations at the 14th patent.

Similar reasoning could be applied to the assignees whose *h*-core centroids are located in the 4th quadrant. For example, SEIKO EPSON, TOYOTA, and NIPPON DENSO all have greater *h*-indices and we would expect that their *h*-core centroids are located to the right of NORTEL NETWORKS' *h*-core centroid, which is indeed the case. The reason that their *h*-core centroids are located below NORTEL NETWORKS' *h*-core centroid must be due to that the citation distribution of NORTEL NETWORKS is more skewed compared to those of SEIKO EPSON, TOYOTA, and NIPPON DENSO. We indeed find that NORTEL NETWORKS' higher ranking patents do receive more citations, but NORTEL NETWORKS' patents ranked behind the 30th place receive less citations than the corresponding patents of SEIKO EPSON, TOYOTA, and NIPPON DENSO.

From the observation above, we would expect that most of the assignees whose *h*-core centroids are located in the 2nd and 4th quadrants of the reference point are those whose *h*-core segments interleave with that of the target, and whose *h*-core areas do not have domination relationship with that of the target. To see how these assignees could be differentiated, their respective *h*-core areas are plotted in Fig. 7. For easier observation, the markers are drawn at intervals, the y-axis is

Table 1
Relevant data for the assignees in 2nd and 4th quadrants of Fig. 5.

Assignee	Quadrant	*h*-Index, *n*	Rank by *n*	*h*-Core centroid		A_c
				c_x	c_y	
CANON INC.	2	92	18	23	452	22,371
QUALCOMM	2	78	32	23	188	13,874
ALTERA CORP.	2	58	54	20	90	7514
SANDISK	2	61	49	22	78	7185
SEIKO EPSON	4	66	45	29	48	5899
TOYOTA	4	67	44	30	45	5736
NORTEL NETWORKS	Target	60	51	25	51	5483
NIPPON DENSO	4	63	48	27	46	5390
NISSAN MOTOR	4	60	50	27	40	4546
PIONEER HI-BRED	2	34	78	8	145	3518

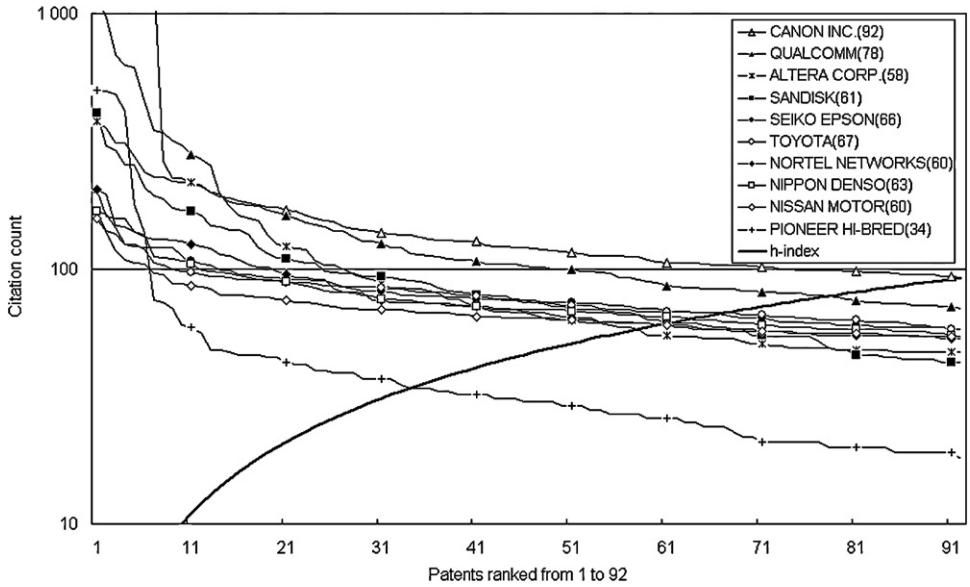

Fig. 7. *h*-Core areas for the assignees in 2nd and 4th quadrants of Fig. 5 (*y*-axis is log-scaled).

log-scaled, citation counts below 10 and over 1 000 are truncated (only CANON INC. and QUALCOMM are affected), and the *h*-indices are connected to indicate the boundaries of the assignees' *h*-core areas.

From Fig. 7, we could easily identify that CANON INC. and QUALCOMM indeed have significantly skewed *h*-core areas and as such their *h*-core centroids fall within the 2nd quadrant. Their *h*-core areas therefore dominate, rather than interleave with, those of NORTEL NETWORKS and the other assignees. The differentiation of CANON INC and QUALCOMM from NORTEL NETWORKS could be conveniently achieved either by comparing their significantly different *h*-indices, or by their *h*-core area sizes (A_c's) (i.e., the numbers of citations received by the *h*-cores), as shown in Table 1 where the assignees are sorted in descending order of their A_c's.

A_c, or *R*-index as A_c is equal to the square of *R*-index, seems to be more desirable. On one hand, A_c indeed correctly ranks the assignees with significantly different *h*-indices as described above. On the other hand, for the rest of the assignees of Table 1 whose *h*-core areas are severely overlapped, A_c successfully suggest a correct ranking order while the *h*-index fails to do so. For example, ALTERA CORP. has the second smallest *h*-index 58, yet it is correctly ranked at the 3rd place by A_c which could be verified from Fig. 7. As another example which could also be verified from Fig. 7, SANDISK having *h*-index 61 is correctly ranked above assignees such as SEIKO EPSON, TOYOTA, and NIPPON DENSO, all having greater *h*-indices than that of SANDISK.

4.2. Second approach

The first approach described in Section 4.1 could determine a specific position for a target assignee within a group of assignees with respect to their *h*-core performance. By counting those assignees whose *h*-core centroids are in the 1st quadrant, and those assignees whose *h*-core centroids are in the 2nd and 4th quadrants having greater A_c's, we can determine the total number of assignees outperforming the target assignee. Again, taking NORTEL NETWORKS as example, there are 43 assignees whose *h*-core centroids are located in its 1st quadrant and, together with the information revealed in Table 1, there are total 49 assignees outperforming NORTEL NETWORKS. Following a similar process, we can also determine the total number of assignees that is outperformed by the target assignee. Again, for NORTEL NETWORKS, its 3rd quadrant has 46 assignees and, with the 3 additional ones listed in Table 1, it outperforms 49 assignees. It so happens that the position of NORTEL NETWORKS is very close to its rank 51 by *h*-index. This is not always true as, for example, ALTERA CORP. is ranked at the 58th place by *h*-index yet it outperforms NORTEL NETWORK as indicated by Table 1. ALTERA CORP. therefore should have a position much different from its rank by *h*-index. However, the closeness of the positions determined by the first approach and the ranks by *h*-index does conform to the observation by Kuan et al. (2011) that *h*-index indeed offers a coarse ranking capability.

For a large group of assignees, repeating the foregoing process for each assignee is cumbersome. However, we could provide a quick qualitative categorization of these assignees with much reduced effort. In this second approach, the two-dimensional distribution of the *h*-core centroids for a group of assignees is overlaid with an $l \times m$ grid and the *h*-core centroids are as such partitioned into $l \times m$ cells. For the assignees whose *h*-core centroids within the same grid cell, they could be jointly considered as belonging to a same qualitative category.

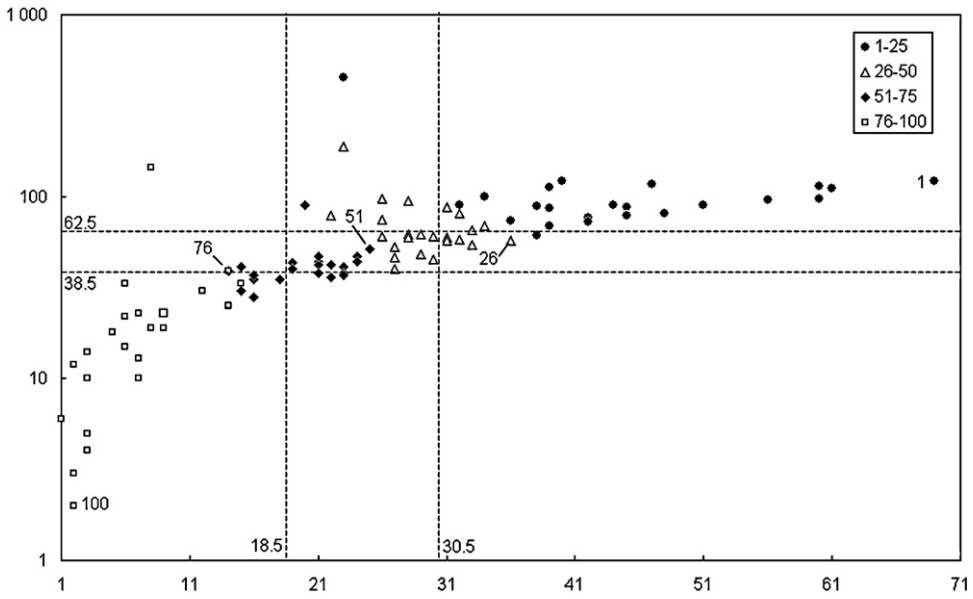

Fig. 8. Partitioning of the *h*-core centroids into 9 grid cells (*y*-axis is log-scaled).

Without losing generality, we choose $l = m = 3$ and the 100 assignees are separated into 9 cells as follows. First, the *h*-core centroids of the 100 assignees are obtained according to Eqs. (3) and (4). Then, by sorting their c_x's, we can easily determine a first *x*-threshold 18.5 and a second *x*-threshold 30.5 where 1/3 assignees (or 33 assignees to be exact) having their $c_x < 18.5$, and another 1/3 assignees (or 34 assignees to be exact) having their $c_x > 30.5$.

Similarly, by sorting the c_y's of these *h*-core centroids, we can also determine a first *y*-threshold 38.5 and a second *y*-threshold 62.5 where 1/3 assignees (or 33 assignees to be exact) having their $c_y < 38.5$, and another 1/3 assignees (or 34 assignees to be exact) having their $c_y > 62.5$.

In other words, the first and second *x*-thresholds together separate the 100 assignees' *h*-core centroids laterally into three non-overlapping groups having substantially the same number of *h*-core centroids, and the first and second *y*-thresholds separate the *h*-core centroids vertically into three non-overlapping groups having substantially the same number of *h*-core centroids, thereby achieving the 3×3 grid. The 9-cell grid is then laid over Fig. 3 and the result is shown in Fig. 8.

As asserted at the end of Section 2, the *x*- and *y*-coordinates of the *h*-core centroid are measures of the productivity and impact sides of an assignee's *h*-core where patents of more citations are more favored. Additionally, the *x*- and *y*-coordinates could jointly provide a measure for the skewness of the *h*-core. Therefore, we can qualitatively describe the assignees whose *h*-core centroids are in a particular grid cell. Table 2 summarizes the qualitative categorization with each table cell corresponding to a grid cell at the same location in Fig. 8. In Table 2, the terms *low*, *medium*, *high* stand for below average, average, and above average performance.

As illustrated in Fig. 8 and Table 2, most assignees are positioned diagonally from the lower left, low-productivity–low-impact category to the upper right, high-productivity–high-impact category, and those positioned in the center, medium-productivity–medium-impact category could be considered the mediocre ones.

Fig. 8 and Table 2 also indicate that, for the 100 assignees, it is quite rare that one has average or above average productivity yet achieves below average impact, as the three grid cells around the lower right corner are either empty or scarcely populated.

As to the categories around the upper left corner, there are few assignees having below average productivity yet with average or above average impact. It is however possible that an assignee having average or above average productivity and above average impact falls within the low-productivity-high-impact category due to the high skewness of its citation distribution.

Table 2
Qualitative categorization of assignees.

Low productivity/ high skewness High impact	Medium productivity/ high skewness High impact	High productivity High impact
Low productivity Medium impact	Medium productivity Medium impact	High productivity Medium impact
Low productivity Low impact	Medium productivity Low impact	High productivity Low impact

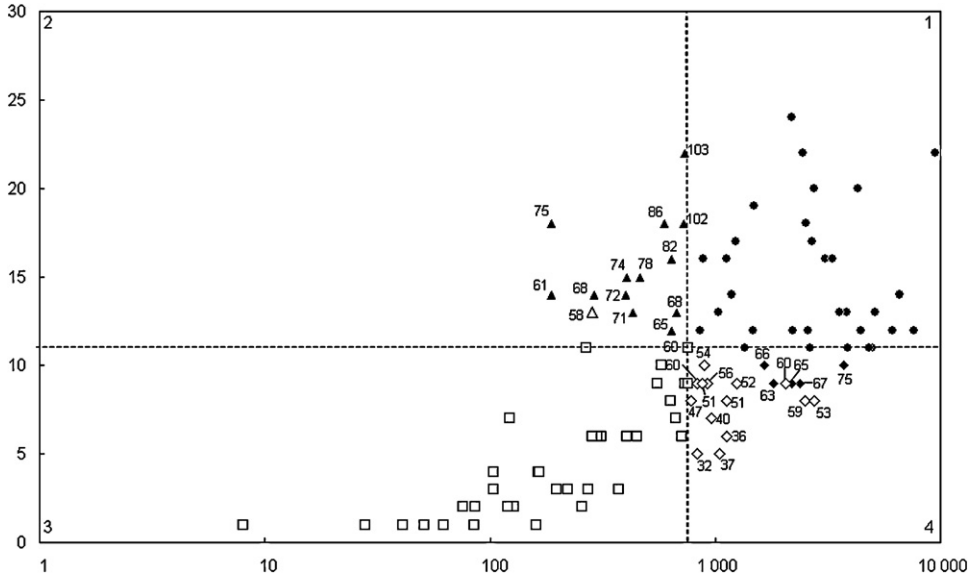

Fig. 9. Distribution of *h*-tail centroids relative to a reference point (*x*-axis is log-scaled).

Similarly, for the more populated medium-productivity–high-impact category, it is also possible that an assignee having above average productivity and impact falls within this category also due to the high skewness of its citation distribution.

With the second approach outlined above, an assignee could be quickly and qualitatively categorized, thereby gaining an immediate understanding of the nature of its performance. This approach also allows us to quickly identify those assignees sharing the same performance nature and those having different natures.

Additionally, we believe that the qualitative categorization could also be valuable in the dynamic analysis of patent assignees. For example, if one assignee is observed to have its *h*-core centroid moving from low-productivity–low-impact category, not diagonally to the medium-productivity–medium-impact category, but vertically up to the low-productivity–medium-impact category, we could conclude that its highly cited patents continue to receive more citations which is of little contribution to boost the assignee's *h*-index.

5. Positioning assignees by *h*-tail centroids

The approaches and the ideas behind the analysis of assignees' *h*-cores described in Sections 4.1 and 4.2 are equally applicable to the analysis of their *h*-tails as well. Without repeating the same analysis process, we will mainly focus on the first approach, and especially on the treatment of the *h*-tail centroids located in the 2nd and 4th quadrants of the reference point.

Again, the assignee NORTEL NETWORKS whose *h*-index 60 is ranked at the 51st place among the 100 assignees is chosen as our target and its *h*-tail centroid is used as the reference point. Then, the *h*-tail centroids of the rest of the assignees are obtained according to Eqs. (5) and (6), and then partitioned relative to the reference point into four quadrants numbered from 1 to 4 at the corners of Fig. 9.

Again, a hollow marker indicates that the *h*-tail centroid's corresponding *h*-index is less than or equal to that of NORTEL NETWORKS, and a solid marker indicates the opposite. Exactly like what is observed in Fig. 5, the *h*-tail centroids in the 1st quadrant all have corresponding *h*-indices greater than NORTEL NETWORKS' *h*-index, and those in the 3rd quadrant all have corresponding *h*-indices less than or equal to NORTEL NETWORKS' *h*-index. As to those located in the 2nd and 4th quadrants, again, both conditions are present and they are labeled by their respective *h*-indices.

As A_c is used to differentiate the assignees whose *h*-core centroids are in the reference point's 2nd and 4th quadrant from the target, it is intuitive to speculate that A_t could play the same role here as well. However, unlike the assignees' *h*-cores, there are a large number of *h*-tail centroids located in the reference point's 2nd and 4th quadrants (14 in the 2nd quadrant and 19 in the 4th quadrant), reflecting the severe variations among the assignees' *h*-tails. For such a large number of assignees, their analysis is very difficult to present in an orderly and readable manner. In the following we therefore describe only the result from the analysis of the fewer assignees whose *h*-tail centroids are in the 2nd quadrant, and the relevant data of these assignees is summarized in Table 3. Even though their analysis is not presented here, our observation is applicable to the assignees whose *h*-tail centroids are located in the 4th quadrant as well.

For the first four assignees listed in Table 3, they have much greater *h*-indices than NORTEL NETWORKS yet their *h*-tail centroids are located slightly to the left of NORTEL NETWORKS' *h*-tail centroid. This scenario must be resulted from their

Table 3
Relevant data for the assignees in 2nd quadrant of Fig. 9.

Assignee	Quadrant	h-Index, n	Rank by n	h-Tail centroid		A_t
				t_x	t_y	
MEDTRONIC INC.	2	103	12	737	22	59,111
SEMICONDUCTOR ENERGY LAB.	2	102	13	729	18	43,014
APPLE	2	86	24	594	18	36,671
TOKYO ELECTRON	2	82	27	638	16	32,637
NORTEL NETWORKS	Target	60	51	758	11	30,931
SCHLUMBERGER	2	68	42	675	13	30,376
CISCO	2	65	47	640	12	24,680
QUALCOMM	2	78	32	466	15	21,736
HALLIBURTON	2	74	36	404	15	18,488
XILINX	2	72	39	403	14	18,178
BAKER HUGHES	2	71	40	433	13	18,092
EMC	2	68	43	291	14	12,539
ALTERA CORP.	2	58	54	285	13	11,507
BOSTON SCIENTIFIC	2	75	34	188	18	8401
SANDISK	2	61	49	187	14	6786

significantly skewed h-tails, as already explained in Section 4.1. This is indeed the case after examining their respective data.

To see whether the rest 10 assignees could be successfully differentiated from NORTEL NETWORKS, their respective h-tail areas are depicted in Fig. 10. For easier observation, the markers are drawn at intervals, the x-axis is log-scaled, and only the patents ranked after the 100th place are shown.

By closely examining Fig. 10, we can see that NORTEL NETWORKS indeed outperforms the 10 assignees as correctly suggested by its greater A_t. Even though for the first 1,000 patents in its h-tails, NORTEL NETWORKS' h-tail segment runs somewhat below that of SCHLUMBERGER, it begins to surface above SCHLUMBERGER and extend farther to the right. In other words, NORTEL NETWORKS receives equal to more citations than SCHLUMBERGER does shortly after the 1,000th patent (1,006th patent to be exact), and NORTEL NETWORKS has greater N_c (3,030 cited patents) than that of SCHLUMBERGER (2,568 cited patents).

As in Section 4.1, we could determine the total number of assignees outperforming or being outperformed by NORTEL NETWORKS with respect to their h-tail performance by counting those assignees whose h-tail centroids are in the 1st or 3rd quadrant, and those assignees whose h-tail centroids are in the 2nd and 4th quadrants having greater or smaller A_t's.

The distribution of the h-tail centroids could also be qualitatively categorized as in Section 4.2. Instead of providing an additional diagram, we could imagine that Fig. 9 is overlaid with a 2×2 grid, which turns out to be quite close to the reality. The exact x-threshold 758 is precisely the same as NORTEL NETWORKS' t_x, and the y-threshold 10 is right next to NORTEL

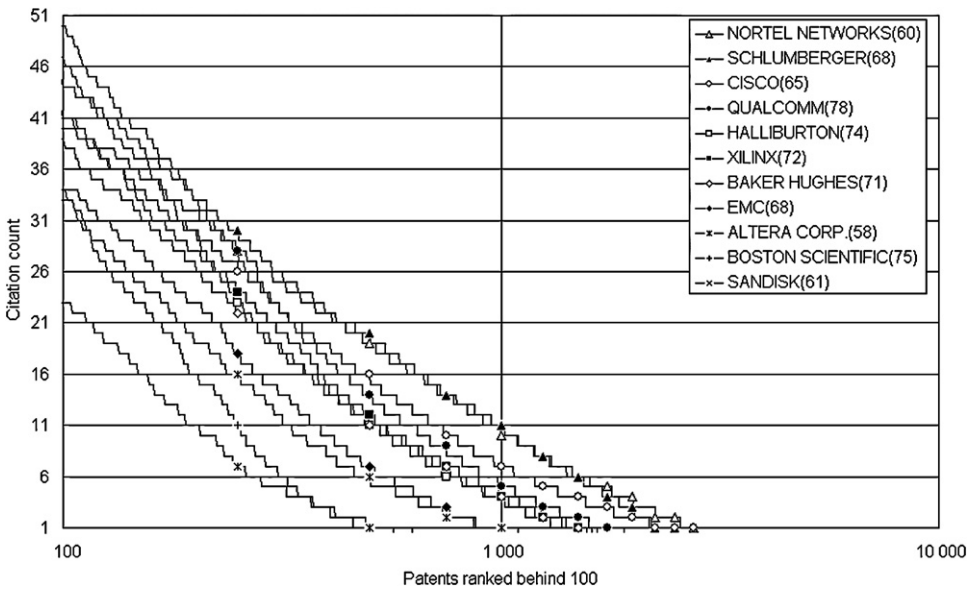

Fig. 10. h-Tail areas for the assignees in 2nd quadrant of Fig. 9 (x-axis is log-scaled).

NETWORKS' t_y 11. Then, the assignees in the four quadrants (or cells) could be categorized as being high-productivity–high-impact (1st quadrant), low-productivity–high-impact (2nd quadrant), low-productivity–low-impact (3rd quadrant), and high-productivity–low-impact (4th quadrant) ones with respect to their *h*-tails.

Similarly, a sizable number of assignees are positioned diagonally from the lower left, low-productivity–low-impact category to the upper right, high-productivity–high-impact category. It is also possible that an assignee having above average productivity and impact falls within the low-productivity–high-impact category due to the high skewness of its citation distribution, as suggested by the four assignees, MEDTRONIC INC., SEMICONDUCTOR ENERGY LAB., APPLE, and TOKYO ELECTRON. This simplified example also reveals that 2×2 grid is too coarse and a 3×3 or finer grid should be more appropriate for our qualitative categorization.

6. Conclusion

Believing that the shapes, rather than the various area sizes, of the *h*-core and *h*-tail areas of an individual more accurately reflect the individual's *h*-core and *h*-tail performance, and that the centroids of the *h*-core and *h*-tail areas faithfully characterize their shapes, we propose to use the *h*-core and *h*-tail centroids as characteristic points and, by plotting these characteristic points in two-dimensional coordinate systems, these individuals' performance with respect to their *h*-cores and *h*-tails could be immediately and conveniently positioned relative to each other.

Our methodology is proven by empirical patent assignee data to be capable of determining those assignees outperforming a target assignee or those assignees outperformed by the target assignee. Additionally, we propose a quick qualitative categorization of a large group of assignees into $l \times m$ categories by overlaying an $l \times m$ grid over the distribution of *h*-core and *h*-tail centroids where the grid lines are dynamically determined by the centroid data of these assignees. With this approach, we can quickly gain an overview of the natures of these assignees' performance.

In addition to the static analysis described by this paper, our approaches are believed to be valuable to the dynamic analysis of performance evolution over a period of time. By observing how the *x*- and *y*-coordinates of an assignee's *h*-core and *h*-tail centroids vary, and how the assignee progresses from one qualitative category to another, we can gain significant insight into the possible sources of change.

Our approaches are also flexible as we can choose to evaluate and compare assignees' relative performance only with respect to their *h*-cores, or only with respect to their *h*-tails, or with respect to both.

Even though the paper uses patent assignees as an illustrative case, the methodology, observations, and results are believed to be equally applicable to individuals at various levels for the evaluation of their research or innovation performance in terms of their publications or patents. However, when the methodology is applied to research performance evaluation, we believe that *h*-core centroids are of greater importance as they are able to obviate the shortcoming of the *h*-index's insensitivity to exceptionally highly cited papers, an issue of great concern to scientific community. On the other hand, when the methodology is applied to innovation performance evaluation, the analysis to the *h*-tail centroids is more important as a great majority of patents are either un-cited or lowly cited.

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.

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*, 577–588.

Bornmann, L., & Daniel, H.-D. (2009). The state of *h* index research, is the *h* index the ideal way to measure research performance? *EMBO Reports*, *10*, 2–6.

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. (2010a). Characteristic scores and scales based on *h*-type indices. *Journal of Informetrics*, *4*, 14–22.

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

García-Pérez, M. A. (2009). A multidimensional extension to Hirsch's *h*-index. *Scientometrics*, *81*, 779–785.

Glänzel, W., & Schubert, A. (1988). Characteristic scores and scales in assessing citation impact. *Journal of Information Science*, *14*, 123–127.

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.

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

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., 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. (2011). Ranking patent assignee performance by *h*-index and shape descriptors. *Journal of Informetrics*, *5*, 303–312.

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

Thor, A., & Bornmann, L. (n.d.). Web application to calculate the single publication *h* index (and further metrics) based on Google Scholar. In *Database Group Leipzig*. Retrieved from http://labs.dbs.unileipzig.de/gsh/

U.S. Patent and Trademark Office. (2010). *Patenting by Organizations 2009*. A Patent Technology Monitoring Team Report, U.S. Patent and Trademark Office. Available from: http://www.uspto.gov/web/offices/ac/ido/oeip/taf/topo_09.htm
van Eck, N. J., & Waltman, L. (2008). Generalizing the *h*-and *g*-indices. *Journal of Informetrics*, *2*, 263–271.
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., & Rousseau, R. (2010). Probing the *h*-core: An investigation of the tail-core ratio for rank distributions. *Scientometrics*, *84*, 431–439.
Ye, F. Y. (2010). Two *h*-mixed synthetic indices for the assessment of research performance. *Journal of Library and Information Studies*, *8*, 1–9.
Zhang, C.-T. (2009). The *e*-index, complementing the *h*-index for excess citations. *PLoS ONE*, *4*, e5429.