Research on the Relationship Between H Index Ranking Evaluation Method and Comprehensive Ranking Based on Big Data Statistics

Shufen Yang*
East China Jiaotong University, Nanchang, China

*Corresponding author: shufenyang@ecjtu.jx.cn

Abstract. The h index of universities can quantitatively calculate the scale of high-level scientific research personnel in universities, and evaluate the academic influence of universities from a new perspective. It has the advantages of simplicity, intuition, and visibility, but there are also some problems. Based on this research background, the paper uses big data statistical methods to conduct an in-depth theoretical analysis of the problems existing in the h-index of universities. The results show that the h index and other ranking indexes are effective in the academic evaluation of colleges and universities. And the correlation analysis between their comprehensive rankings and various evaluation indicators shows that the correlation between the rankings of various indexes and the comprehensive rankings is higher than that of SCI included rankings.

Keywords: Big data statistics, university H index, ranking, comprehensive ranking, rank relationship.

1. Introduction
In 2005, the American physicist Professor Hirsch proposed a new evaluation index, the h index, which organically combines the quantity and quality of the papers, while considering the quantity and quality of papers, rather than simply using the number of papers or the frequency of citations to evaluate scientists Personal academic influence is more reasonable [1]. As soon as the h index was put forward, it received widespread attention and was quickly promoted to the evaluation and research of institutions, journals, patents and other fields. Based on this, this paper further analyses the problems of institution h index used in institution evaluation, and proposes corresponding countermeasures against the impact of cooperative publications on institution h index and the low degree of distinction of h index, and puts forward corresponding countermeasures. Based on h, a comprehensive evaluation method for the academic influence of institutions is obtained using factor analysis.

2. The application of H index in the analysis of university influence

2.1. Concept elaboration
In the analysis of the influence of colleges and universities, drawing on and transplanting relevant indicators of information metrology is conducive to expanding research ideas and methods. Taking
How Net as an example, some scholars have studied the h-index of academic publications and believe that the h-index is a useful tool for the evaluation of academic influence [2]. The concept of h index was proposed by Hirsch and modified by other scholars. It is defined as: the papers published by scientist S are arranged in descending order of the number of citations, and papers with the same number of citations have different serial numbers, if and only the current papers are h When the number of citations for each paper is at least h, and the number of citations of the h+1th paper is less than h+1, the h index of scientist S is h.

Through the empirical application of the h index in evaluation research, its characteristics can be summarized as follows: the calculation is relatively simple; the h index takes into account the quality and quantity of academic results; the results are relatively stable and accurate; it has data source sensitivity and evaluation fields Sensitivity, the data source and the evaluation field are different, and the results obtained are also different; it has the sensitivity of the measurement time interval, etc. The h index can be used in the evaluation and analysis of corresponding information sources [3]. It was initially used to evaluate the academic achievements of scientists, and then extended and expanded by many researchers to be applied to the evaluation of academic journals, the evaluation of academic institutions such as universities, and the evaluation of patents.

2.2. The H index is used to evaluate the influence of academic achievements

University academic achievements cited H index: The following methods can be used to obtain the H index of an academic achievement user: First, obtain the number of citations of all academic achievements of the user, sort them in descending order of the number of citations, and indicate the serial number; then, find out which serial number is not If it is greater than the corresponding maximum number of citations, this number is the H index of the user. This article takes the publication of academic achievements of a university from 2019.11.29 to 2019.12.5 on CNKI as an example. The number of citations for the academic achievement of serial number 35 is 35. At this time, the number of citations for the academic achievement of serial number 36 is 33, which is less than its serial number 36. Therefore, the H index of the academic achievement of a university is 35.

The calculation of the university G index: The papers published by each university are arranged in descending order of the number of citations, and the sum of the cumulative citations of the first g highly cited papers published by each university is higher than or equal to g2.

The calculation of the W index of colleges and universities: the papers published by each college are arranged in descending order of the number of citations. The w value of a college is equal to if and only if there are w papers in its published papers, and each paper has been obtained no less than 10w times. The number of citations for each of the remaining papers is less than 10 (w+1).

The calculation of the A index of each university: The papers published by each university are ranked in descending order of the number of citations, and the average value of the citations of the papers in the performance core delineated by the H index.

The calculation of the university R index: the papers published by each university are arranged in descending order of the number of citations, and the square root of the sum of the citation numbers of the papers in the performance core delineated by the H index. The calculation model for the overall ranking of the comprehensive index is as follows:

\[ H = lG + mW + nH + kA + pR \]  

\[ l + m + n + k + p = 1 \]
According to the above definitions of h index, G index, W index, A index, and R index, the results are shown in Table 1, which lists each index value of each university and the corresponding index ranking.

| University                                      | h        | G        | W        | R        | A        | Overall ranking |
|------------------------------------------------|----------|----------|----------|----------|----------|-----------------|
| Tsinghua University                            | 148(2)   | 212(2)   | 29(2)    | 190.39(2)| 244.93(2)| 2               |
| Beijing University                             | 150(1)   | 229(1)   | 32(1)    | 204.33(1)| 278.35(1)| 1               |
| Zhejiang University                            | 104(8)   | 154(8)   | 19(9)    | 138.37(8)| 184.11(7)| 8               |
| Nanjing University                             | 107(7)   | 153(9)   | 21(6)    | 137.49(9)| 176.67(10)| 7               |
| Shanghai Jiaotong University                   | 110(5)   | 155(6)   | 22(5)    | 140.61(6)| 179.74(9)| 5               |
| Sun Yat-sen University                         | 109(6)   | 155(6)   | 21(6)    | 140.04(7)| 179.93(8)| 6               |
| Fudan University                               | 114(4)   | 163(5)   | 21(6)    | 146.8(5)| 189.06(6)| 4               |
| University of Science and Technology of China   | 119(3)   | 178(3)   | 24(4)    | 158.45(3)| 210.97(4)| 3               |
| Shan Dong University                            | 76(13)   | 104(17)  | 14(17)   | 94.75(17)| 118.13(18)| 13              |
| Wuhan University                               | 85(12)   | 123(13)  | 16(12)   | 110.83(13)| 144.52(12)| 12              |
| Sichuan University                             | 74(15)   | 104(17)  | 14(18)   | 93.89(18)| 117.53(19)| 18              |
| Jilin University                               | 91(10)   | 131(10)  | 19(9)    | 118.03(10)| 153.08(11)| 10              |
| Huazhong University of Science and Technology  | 70(18)   | 96(20)   | 13(19)   | 86.73(20)| 107.46(24)| 19              |
| Harbin Institute of Technology                 | 66(22)   | 93(23)   | 12(24)   | 83.69(23)| 106.12(26)| 23              |
| Nankai University                              | 87(11)   | 124(12)  | 17(11)   | 111.53(12)| 142.98(13)| 14              |
| Xian Jiaotong University                       | 64(23)   | 94(22)   | 11(26)   | 84.9(22)| 112.63(21)| 25              |
| Northeastern University                        | 104(8)   | 174(4)   | 25(3)    | 155.55(4)| 230.45(3)| 9                |
| Tianjin University                             | 64(23)   | 92(24)   | 13(19)   | 82.69(26)| 106.84(25)| 22              |
| Dalian University of Technology                | 73(17)   | 107(16)  | 15(14)   | 96.03(16)| 126.32(17)| 17              |
| Central South University                       | 61(26)   | 92(24)   | 13(19)   | 82.92(25)| 112.7(20)| 21              |
| Lanzhou University                             | 67(20)   | 88(27)   | 11(26)   | 80.34(27)| 96.34(28)| 27              |
| Southeast University                           | 74(16)   | 108(15)  | 16(12)   | 97.5(15)| 128.46(16)| 15              |
| Beijing Normal University                      | 70(18)   | 131(10)  | 15(14)   | 117.33(11)| 196.67(5)| 11              |
| Xiamen University                              | 75(14)   | 113(14)  | 13(14)   | 101.07(14)| 136.21(14)| 16              |
| Tongji University                              | 59(27)   | 98(19)   | 13(19)   | 87.83(19)| 130.75(15)| 20              |
| South China University of Technology           | 63(25)   | 92(24)   | 13(19)   | 83.04(24)| 109.44(23)| 26              |
| Hunan University                               | 67(20)   | 96(20)   | 12(24)   | 86.27(21)| 111.07(22)| 24              |
| Chongqing University                           | 59(27)   | 85(28)   | 11(26)   | 76.75(28)| 99.83(27)| 28              |
| Northwestern Polytechnical University           | 39(33)   | 53(33)   | 7(34)    | 47.17(33)| 55.63(34)| 33              |
| Beijing University of Aeronautics and Astronauts| 51(29)   | 77(30)   | 11(26)   | 69.12(30)| 93.67(30)| 29              |
| Beijing Institute of Technology                | 51(30)   | 78(29)   | 10(30)   | 69.86(29)| 95.71(29)| 30              |
| University of Electronic Science and Technology| 41(32)   | 62(32)   | 9(31)    | 55.68(32)| 75.61(32)| 32              |
| Ocean University of China                      | 48(31)   | 70(31)   | 9(31)    | 63.08(31)| 82.9(31)| 31              |
| Renmin University of China                     | 34(34)   | 51(34)   | 8(33)    | 46.45(34)| 63.47(33)| 34              |

As shown in Table 1, the total ranking is the combination of the various indexes, and the indexes are sorted separately, and then according to the 6 indexes and SCI rankings, which universities are in
the top 10, and 6 indexes are among the top 10 universities. Which ones, and so on, 7 items are better than 6 items, 6 items are better than 5 items, etc.; if the number of indicators entering the top 10 is equal, the sum of the rankings of which university is smaller, and finally get the ordinal number of an overall ranking [4]. According to the above ranking principle, the top 10 universities are Peking University, Tsinghua University, University of Science and Technology of China, Fudan University, Shanghai Jiaotong University, Sun Yat-sen University, Zhejiang University, Nanjing University, Northeastern University and Jilin University.

3. Result analysis

Before performing correlation analysis, it is necessary to perform a single-sample Kolmogorov-Smirn test. The KS test is used to test whether the distribution of the sample obeys a certain theoretical distribution. The theoretical distribution here can be normal [5]. Distribution, uniform distribution, Poisson distribution and exponential distribution. Through the K-S test on the ranking ordinal numbers of these 8 groups, the significance probability is all > 0.05. Therefore, the data distribution of these 8 groups does obey the normal distribution, as shown in Figure 1.

![Figure 1. A single sample of data obeys a normal distribution](image)

Correlation analysis usually uses the sample correlation coefficient $r$ to measure the degree of correlation between variables, $-1 \leq r \leq 1$; the closer $|r|$ is to 1, the higher the degree of linear correlation between variables. First, assume that the overall correlation is zero, that is, $H_0$ is that there is no significant linear correlation between the two populations [6]. If the associated probability is less than or equal to the specified significance level, then $H_0$ is rejected and the two populations are considered to have a significant linear correlation; otherwise, there is no significant linear correlation. Significant linear correlation.

The correlation coefficients of "985" first-phase university h index ranking and Wushulian ranking, Wangda ranking and Wuhan University ranking are 0.608, 0.563 and 0.617 respectively, and the associated probability is less than 0.01. It can be seen that under the condition of a significance level of 0.01, there is a significant positive correlation between the ranking of the "985" university h index and the ranking of each university [7]. The correlation coefficients between Wushulian ranking and Wangshulian ranking, Wushulian ranking and Wuhan University ranking, and Wangshulian ranking and Wuhan University ranking are respectively 0.877, 0.935, and 0.924, and the associated probabilities are all less than 0.01. It can be seen that under the condition of a significance level of 0.01, there is a very significant positive correlation between the three university rankings. Figure 2 shows the positive correlation between h index rankings and university rankings.
Figure 2. Positive correlation between h index ranking and university ranking

4. Discussion

4.1. The value of the h index is related to the conditions under which it is based
The value of the h index is related to the database based on, the age of the cited literature, and whether it contains self-citation and other conditions. The h index increases as the number of years of citation increases. This paper does not exclude self-citation when calculating the h index. It can be seen from Figure 3 that the correlation between h and G is not very high, and the correlation coefficient is only 0.57. This may be related to the submission orientation of major universities and the selection of journals included by ISI. Except for a few singularities, the values of h and W are quite close.

Figure 3. Correlation between h and G

4.2. The relationship between the h index and other measurement indicators
R’s citation report provides not only h, but also the number of articles R and the total cited R of these articles. The correlation between h and the amount of papers R and the total citations R are shown in Figure 4. The correlation coefficients are 0.81 and 0.93, respectively. The average value of the total cited R in Table 1 is about 5 times the square of the average value of h, which is basically consistent with Hirsch's statement. Although the correlation between h and the amount of papers R and the total cited R is relatively good, the three indicators have different meanings and are independent of each other. In his paper, Hirsch analysed the advantages and disadvantages of these three indicators and the relationship between them. The h index uses a simple and clear method to combine the indicators that indicate quantity (the number of papers) and the indicators that indicate quality (total citations), which
has a certain comprehensiveness. R lists these three indicators together in its citation report, which can provide a more complete description of the scientific achievements of an institution.

Figure 4. Correlation between h and the amount of paper R

5. Conclusion
This paper studies the relationship between the h index of my country's "985" universities and the comprehensive university rankings from the perspective of bibliometrics, and finds that the h index is simple and easy to obtain, and is suitable for the comprehensive evaluation of university research output and influence; h index ranking and the comprehensive university rankings are correlated and are a useful reference index for university rankings.

References
[1] Meho, L. I., & Rogers, Y. Citation counting, citation ranking, and h-index of human-computer interaction researchers: a comparison of Scopus and Web of Science. Journal of the American Society for Information Science and Technology, 59(11) (2008) 1711-1726.
[2] Harzing, A. W., & Van der Wal, R. A Google Scholar h-index for journals: An alternative metric to measure journal impact in economics and business. Journal of the American Society for Information Science and technology, 60(1) (2009) 41-46.
[3] Lopez, J., Susarla, S. M., Swanson, E. W., Calotta, N., & Lifchez, S. D. The association of the H-index and academic rank among full-time academic hand surgeons affiliated with fellowship programs. The Journal of hand surgery, 40(7) (2015)1434-1441.
[4] Serenko, A., & Bontis, N. A citation-based ranking of the business ethics scholarly journals. International Journal of Business Governance and Ethics, 4(4) (2009) 390-399.
[5] Alonso, S., Cabrerizo, F. J., Herrera-Viedma, E., & Herrera, F. h-Index: A review focused in its variants, computation and standardization for different scientific fields. Journal of informetrics, 3(4) (2009) 273-289.
[6] Mingers, J., & Yang, L. Evaluating journal quality: A review of journal citation indicators and ranking in business and management. European Journal of Operational Research, 257(1) (2017) 323-337.
[7] Sarkiss, C. A., Riley, K. J., Hernandez, C. M., Oermann, E. K., Ladner, T. R., Bederson, J. B., & Shrivastava, R. K. Academic productivity of US neurosurgery residents as measured by H-index: program ranking with correlation to faculty productivity. Neurosurgery, 80(6) (2017) 975-984.