Scientometrics (2018) 117:1721–1732
https://doi.org/10.1007/s11192-018-2918-0

The representative works of scientists

Jianlin Zhou[1] · An Zeng[1] · Ying Fan[1] · Zengru Di[1]

Received: 14 March 2018 / Published online: 5 October 2018
© Akadémiai Kiadó, Budapest, Hungary 2018

Abstract

Nowadays identifying the personal representative works is becoming increasingly important and necessary for scientists in many cases, such as faculty hiring and promotion applications. There are already a few methods based on different criteria for selecting the representative works of a scientist, like citation count. In addition, we can observe that some researchers always produce many similar quality scientific papers and some researchers have several highly cited papers compared with his or her other papers. In this context, we propose to use the maximum gap in a histogram of a scientist's sorted papers' citation counts to classify his or her papers into two groups, i.e. representative papers and regular papers. Based on the maximum gap, we then design an indicator D_r to quantify the impact difference between scientist's representative works and regular works. We apply this selection method and D_r index into the data of American Physical Society (APS) journals. The results indicate that the selection method can better identify the representative works of Nobel laureates in Physics compared with using the most cited paper. We also find that the number of representative works selected by our method is related to D_r. A larger number of selected papers would appear when the value of D_r index is relatively smaller. Meanwhile, we also observe that D_r is weakly correlated with the h index and total citation.

Keywords Representative work · Citation count · Maximum gap

Introduction

The ongoing rapid development of information technologies has greatly accelerated the publication of scientific findings, resulting in a large number of scientific papers (Larsen and Von Ins 2010). The quantitative studies of these papers thus become a major way to evaluate the scientific influence of researchers (Sinatra et al. 2016). As many existing metrics are in scale with the number of papers, the measured influence of a researcher

✉ An Zeng
anzeng@bnu.edu.cn

✉ Ying Fan
yfan@bnu.edu.cn

[1] School of Systems Science, Beijing Normal University, Beijing 100875, People's Republic of China

depends more on the number of published papers instead of the quality of papers (Medo and Cimini 2016). To remove the size effect of the traditional metrics, nowadays researchers are requested to list a limited number of their representative papers (or called selected papers) when applying for a grant or promotion (Clauset et al. 2015).

In many occasions e.g. hiring, funding and promotion, researchers need to select their representative works. How to select the representative works of a scientist is actually an open question. However, it is generally believed that the representative works of a scientist are superior in quality to his or her other papers. There have been various evaluation methods used to select the representative works of a researcher. For example, people can choose his or her most highly cited paper or the papers published in top journals as their own representative works (Ioannidis et al. 2014). Niu et al. (2016) consider that the representative work of a researcher should be an important paper in his or her area of expertise and propose a self-avoiding preferential diffusion process to identify individual representative works. Considering the different contribution of coauthors in a paper (Bao and Zhai 2017), Bao and Wang (2018) identify the representative works of scientists based on credit allocation. These selection methods measure the quality of the selected paper based on different perspectives or criteria.

Usually a researcher is not to invest his or her effort evenly in each research work and produce many papers with similar quality, but to focus on some promising projects which may result in some representative papers that are better than his or her other papers (Bergstrom et al. 2016; Foster et al. 2015). For a researcher, his or her representative works may be distinct from or close to his or her other papers in quality. The difference between the representative works and regular works of a scientist may be related to the possible research pattern. Although there are several methods applied to select the representative works of a scientist, little attention is paid to measure the difference between the representative papers and regular papers of a scientist. In this work, we use the simplest citation count to evaluate the quality of the paper and calculate the relative large gap compared with other gaps in a histogram of his or her sorted papers' citation counts to measure this difference.

In this paper, firstly we propose a simple method based on citation count to select the representative works of a scientist. We choose the maximum gap in a histogram of his or her sorted papers' citation counts as the dividing line. The papers on the left side of the maximum gap are regarded as the representative works of a researcher. Our selection method can partially identify the representative works of Nobel laureates in Physics. Then we develop a simple metric D_r to quantify the difference between the representative works and regular papers. The difference is actually defined as the size of the gap between the top-cited articles of an author and the other articles of the same author. A larger maximum gap compared with other gaps in a histogram of his or her sorted papers' citation counts indicates that the author has one or several papers that are substantially more influential than his or her other papers. We apply the selection method and indicator D_r into the data from all journals of the American Physical Society (APS). We find that the number of representative works of scientists is related to the D_r index. When the D_r of a scientist is smaller, the number of representative works selected may be more. The total citation and h index are as two important indicators to evaluate the scientific output of a researcher. We also investigate the correlations between the metric D_r and citation, h index (Hirsch 2005). We find that the metric D_r is weakly correlated with the well-known h index and total citation.

Method

Data collection

In this paper, we make use of the publication data from all the journals of American Physical Society which involve Physical Review A, B, C, D, E, I, L, ST and Reviews of Modern Physics. The data contains over 450,000 papers, ranging from year 1893 to year 2009. For each paper, we obtain the information of its title, DOI, author names, affiliations, printed time, received time, references, PACS codes and so on. To reduce the effect of the author name disambiguation, in this paper we make use of the dataset in which Sinatra et al. have conducted a comprehensive disambiguation process in the APS data and a total number of 236,884 distinct authors are identified (Sinatra et al. 2016).

The selection of representative works

For evaluating the quality of publications, many methods have been proposed (Zeng et al. 2017). The simplest method is the citation count (Garfield 1955), which reveals the quality of papers based on their number of citations. Despite that this method just considers the number of citations and ignores these citations' quality, the quality of a paper could be reflected by its citation number (Radicchi et al. 2008; Fiala et al. 2015; Mariani et al. 2016). In this paper, we measure the difference in quality of the two papers by the gap between their citation numbers.

For a researcher, assuming he or she has N papers, the citations of the N papers which are sorted by their citation counts in descending order are denoted by C_1, $C2, \ldots, C_N$ and $C_1 \geq C2 \geq \cdots \geq C_N$. Then the gaps G_1, $G_2, \ldots, G_{N-1}$ are defined as $C_1 - C_2$, $C_2 - C_3, \ldots, C_{N-1} - C_N$, respectively. If the gap G_M is the biggest gap of the $N-1$ gaps, the first M papers are selected as the representative papers of a researcher and the remaining $N-M$ papers are regular papers. In this paper, we use the above method to select the representative works of a scientist. This selection method is simple and it just needs the information of papers' citation counts.

D_r index

In Fig. 1, we show the results of two researchers. Both researchers publish 10 papers, with h index equal to 10. Based on our selection method, the researcher A has one representative work and the researcher B has three representative works. For researcher A, the maximum gap is much bigger than other gaps and the selected representative work has a much higher citation count, which is distinct from other nine papers. However, for researcher B, the maximum gap is not substantially bigger than other gaps, which shows that the difference between representative works and regular works is not big.

To quantify the difference between representative works and regular papers based on the maximum gap, we design a simple indicator D_r. For a researcher, his or her D_r can be expressed as follows,

$$D_r = \max_{1 \leq i \leq n-1} \left\{ \frac{C_i - C_{i+1}}{\max_j(C_j) - \min_j(C_j)} \right\} \left(1 - \frac{1}{\log\left(\sum_j C_j + 10\right)} \right), \tag{1}$$

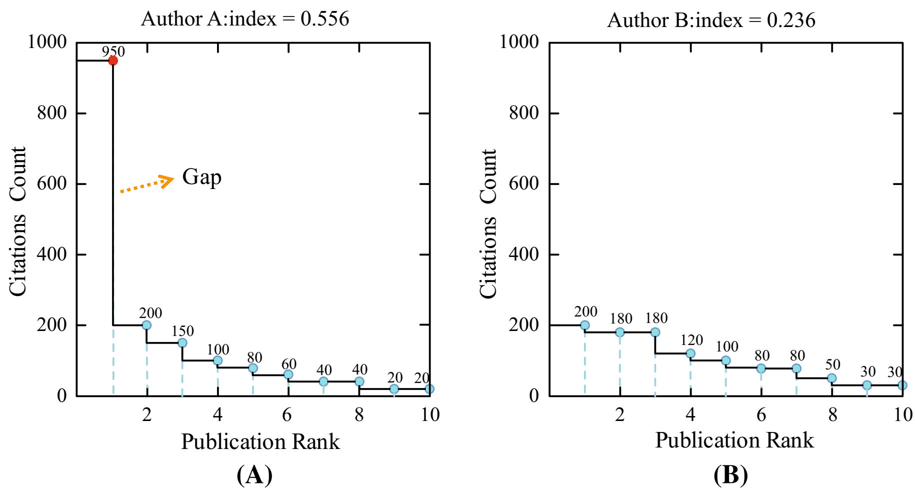

Fig. 1 The illustration of the gap in a histogram of his or her sorted papers' citation counts (i.e. papers' citation versus papers' rank in descending order). Both author A and author B publish 10 papers. As the citations of these papers are over 10, both of these two researchers have h index equal to 10. However, researcher A has a bigger gap in a histogram of sorted papers' citation counts. The D_r value of researcher A is larger than that of researcher B (Color figure online)

where n is the number of his or her papers, C_j is the number of citations for j-th paper which is sorted in descending order by the number of citations, $\max_j(C_j)$ is the maximum number of citations for n papers and $\min_j(C_j)$ is the minimum number of citations for n papers. The D_r index can be regarded as consisting of two parts: D_1 and D_2. Their definitions are as follows:

$$D_1 = \max_{1 \le i \le n-1} \left\{ \frac{C_i - C_{i+1}}{\max_j(C_j) - \min_j(C_j)} \right\}, \quad D_2 = 1 - \frac{1}{\log\left(\sum_j C_j + 10\right)} \tag{2}$$

The D_1 measures the normalized gap between two papers and D_2 is used to remove noise. In APS data, there are numerous scientists whose papers are lowly cited. For example, if a scientist has only one paper with 2 citations while all his or her remaining papers have only 1 citation. Then the first term of D_r index is very large. Clearly, we cannot say that the maximum gap is big as the large relative gap between citation 2 and citation 1 is fake. To suppress the tendency that the scientists with large D_1 index are dominated by the lowly-cited ones, we use D_2 in D_r index to correct the bias. The metric D_r is kept in the range of [0, 1). If D_r is equal to 0, we consider that the papers of a scientist are all similar and we randomly select a paper as his or her representative work. As long as the D_r index is not equal to 0, our method can assign his or her representative works. when the D_r is close to 0, there is almost no impact difference between the selected representative works and regular papers of a scientist qualitatively. When the D_r index gets closer to 1, the selected representative works are obviously distinct from regular papers. When D_r is used in Fig. 1, one can easily see the difference between the two example researchers: researcher A has $D_r = 0.556$ while researcher B has $D_r = 0.236$. It indicates that the difference between the representative works and regular papers of researcher A is bigger than researcher B's.

Empirical results

It is well known that Nobel laureates usually have the representative works which are the so-called Nobel prize-winning papers. The official website of the Nobel Prize provides the summary statement naming the research achievement or discovery for which a Nobel Prize was awarded. With this statement, it is possible to identify the Nobel prize-winning papers best reflecting on those research achievements or discoveries of Nobel laureates (Schlagberger et al. 2016). Those papers often attract a wide range of attention and obtain a lot of citations. In this paper, we choose 28 Nobel laureates in Physics (1995–2013) in Table 1 whose Nobel prize-winning papers are collected in ref. (Shen and Barabási 2014) as our benchmark to examine the effectiveness of selection method. Noted that during the period from 1995 to 2013, there are in total 43 Prize laureates in Physics, with only 27 of them whose Nobel prize-winning papers are in the APS data we used. Based on the selection method of using the maximum gap, we have selected the representative works of 27 Nobel laureates and the results of the comparison with the Nobel prize-winning papers are shown in Table 1. One can find that our selection method can accurately identify the representative works of 19 scientists among 27 Nobel laureates. When using the most cited paper of a Nobel laureate to identify his or her representative works, it can accurately identify representative works of 16 scientists, which indicates that our selection method can better identify the Nobel prize-winning papers. The citations of some Nobel prize-winning papers (i.e. Phys. Rev. Lett. 84, 3232, Phys. Rev. Lett. 29, 1227) in our dataset is relatively small and not substantially distinct from other papers of their corresponding Nobel Prize winners which may make our method based on citation count invalid.

We apply our selection method into APS dataset and study the characteristics of the selected papers. Firstly, we identify 64,233 scientists whose the number of papers is not fewer than 5 and calculate the number of their representative works. Table 1 shows the distributions of the number of their representative works. The range of the number of their selected papers is $[0, 50]$. It indicates that the representative works selected by our method can be many. We can find that more than half of scientists (52.45%) only have one representative work and there are fewer scientists (5.43%) who have more than 5 representative works (Table 2).

Next we calculate the number of publications and the D_r index of these 64,233 scientists. We investigate the relationship of the number of selected papers with D_r index and the number of publications, respectively. The results are shown in Fig. 2. We find that the number of representative works selected by our method is related to D_r. A larger number of selected papers would appear when the value of D_r index is relatively smaller, as shown in Fig. 2a. When the D_r index is less than 0.15, the large number of selected papers (> 20) may appear. when the D_r index of a researcher is more than 0.2, the number of his or her selected papers will not exceed 10. The indicator D_r quantifies the difference between the representative works and regular works of a scientist. When the D_r of a scientist is very small, the impact of the selected representative works and regular works are similar. In this case, some other criteria might need to be used to further identify the true representative works of a scientist. We also find that some low-yielding scientists have larger numbers of selected papers while the high-yielding scientists do not have a large number of selected papers in Fig. 2b.

We further study the properties of D_r index. We calculate the D_1 and D_r index of each author in APS data. Then we measure the Pearson correlation between each index and publications. Their relations are presented with scatter plots in Fig. 3. We can see that

Table 1 The result of using our method to identify the representative works of the Nobel laureates in Physics

ID	Year	Nobel laureates	Nobel prize-winning paper	Citation	No. of selected papers	Included
1	2013	P. W. Higgs	Phys. Rev. Lett. 13, 508 (1964)	164	2	Yes
2	2013	F. Englert	Phys. Rev. Lett. 13, 321 (1964)	172	1	Yes
3	2012	S. Haroche	Phys. Rev. Lett. 77, 4887 (1996)	309	2	Yes
4	2012	D. J. Wineland	Phys. Rev. Lett. 76, 1796 (1996)	210	1	No
5	2008	Y. Nambu	Phys. Rev. 122, 345 (1961)	859	1	Yes
6	2007	A. Fert	Phys. Rev. Lett. 61, 2472 (1988)	806	1	Yes
7	2007	P. Grünberg	Phys. Rev. Lett. 57, 2442 (1986)	380	2	Yes
8	2005	R. J. Glauber	Phys. Rev. Lett. 10, 84 (1963)	162	1	No
9	2005	J. L. Hall	Phys. Rev. Lett. 84, 5102 (2000)	47	1	No
10	2005	T. W. Hänsch	Phys. Rev. Lett. 84, 3232 (2000)	19	1	No
			Phys. Rev. Lett. 84, 5102 (2000)	47		
11	2004	D. J. Gross	Phys. Rev. Lett. 30, 1343 (1973)	457	1	Yes
12	2004	F. Wilczek	Phys. Rev. Lett. 30, 1343 (1973)	457	1	Yes
13	2004	H. D. Politzer	Phys. Rev. Lett. 30, 1346 (1973)	497	1	Yes
14	2003	A. J. Leggett	Phys. Rev. Lett. 29, 1227 (1972)	32	2	No
15	2002	R. Davis	Phys. Rev. Lett. 20, 1205 (1968)	108	2	Yes
16	2002	M. Koshiba	Phys. Rev. Lett. 58, 1490 (1987)	171	1	Yes
17	2002	R. Giacconi	Phys. Rev. Lett. 9, 439 (1962)	10	1	Yes
18	2001	W. Ketterle	Phys. Rev. Lett. 75, 3969 (1995)	1159	1	Yes
19	1998	R. B. Laughlin	Phys. Rev. Lett. 50, 1395(1983)	936	1	Yes
20	1998	D. C. Tsui	Phys. Rev. Lett. 48, 1559 (1982)	607	1	Yes
21	1998	H. L. Stormer	Phys. Rev. Lett. 48, 1559 (1982)	607	1	Yes
22	1997	S. Chu	Phys. Rev. Lett. 55, 48 (1985)	84	1	No
23	1997	C. Cohen-Tannoudji	Phys. Rev. Lett. 61, 826 (1988)	164	1	Yes
24	1997	W. D. Phillips	Phys. Rev. Lett. 61, 169 (1988)	96	5	Yes
25	1996	D. D. Osheroff	Phys. Rev. Lett. 28, 885 (1972)	57	1	No
26	1996	R. C. Richardson	Phys. Rev. Lett. 28, 885 (1972)	57	1	No
27	1996	D. M. Lee	Phys. Rev. Lett. 28, 885 (1972)	57	3	Yes
28	1995	M. L. Perl	Phys. Rev. Lett. 35, 1489 (1975)	~	2	~

Note: "Citation" means the citation of the Nobel prize-winning paper. T. W. Hänsch has two Nobel prize-winning papers and M. L. Perl's Nobel prize-winning paper (Phys. Rev. Lett. 35, 1489) is not in the dataset we use

some authors with small number of papers have large D_1 in Fig. 3a but small D_r in the Fig. 3b, which indicates that the proposed index is not sensitive to the publication. At the same time, the low Pearson correlation coefficient between D_r and publication also confirms that D_r is weakly correlated with publication. Next we compare the results of D_1 and D_r with the well-known h index and citation count using the scatter plot shown in Fig. 4. To better quantify the results, we again compute the Pearson correlation for each scatter plot. The calculated Pearson correlation coefficients show that D_r is weakly correlated with the h index and citation count. D_1 has a much weaker correlation with the existing metrics which are in scale with the number of papers. In addition, we could see the distribution of

Table 2 The distributions of the number of representative works for scientists whose the number of papers is not less than 5

No. of selected papers	0	1	2	3	4	5	> 5 and ≤ 50
Percent (%)	0.05	52.45	21.01	10.91	6.75	3.4	5.43

Note: "No. of selected papers is 0" means that his or her all papers are similar and any paper can be selected as his or her representative work

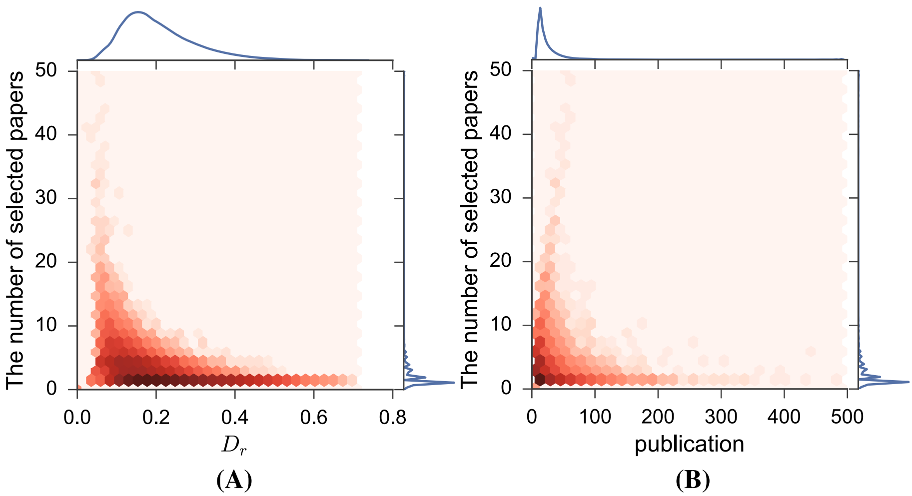

Fig. 2 The scatter plot of researchers' number of selected papers versus their D_r index and total number of publications, respectively. The color here indicates the number of researchers overlapped in each data point. The darker the color is, the more researchers overlap in the data point. In this figure, the distribution of the number of selected papers, D_r, the number of publications are shown as well (Color figure online)

D_r, which exhibits that most researchers have small D_r while some small number of researchers have high D_r.

Based on the above correlation analysis, the D_r index exhibits its own characteristic which is different from h index and citation. The Nobel laureates as a special group, we explore the characteristics of their D_r index. The basic statistical information of the 28 Nobel laureates in Physics is shown in Table 3. We find that D_r indicators of the Nobel Laureates are also different. More than one third of the Nobel Laureates' D_r values are above 0.4 and 5 of 28 Nobel Laureates' D_r values is less than 0.2. Different D_r index can reflect inequalities in the impact of papers of a researcher, which reflects the possible research pattern. The h index is an indicator measuring the scientific impact of a researcher. The D_r index is weakly correlated with h index, indicating that impact value and its inequalities are two different dimensions. Considering the h index as control variable, we aim to reveal the different research patterns adopted by scientists with similar impact. We then calculate the D_r index of each Nobel Laureate and their corresponding different quantiles of D_r index of the scientists with the same h index in Fig. 5. We can observe that more than half of the Nobel Laureates whose D_r exceed the corresponding third quartile of D_r index of the scientists with the same h index. It indicates that the

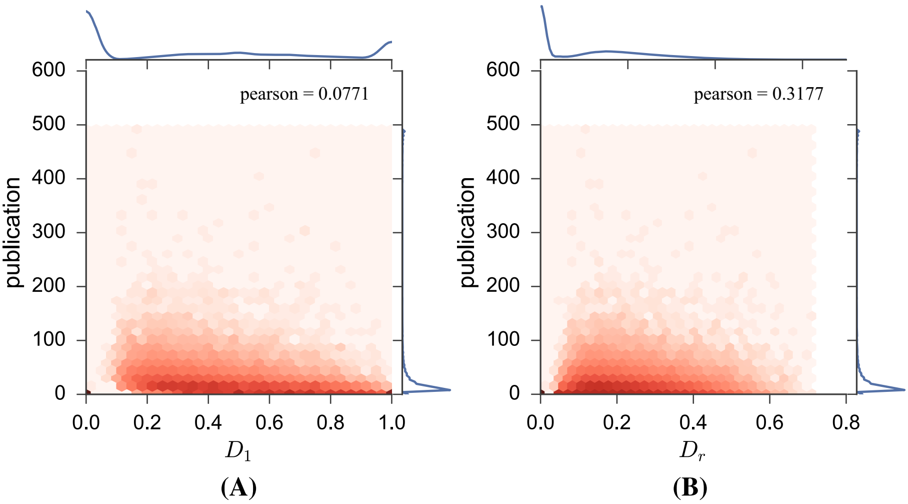

Fig. 3 The scatter plot of researchers' D_1 and D_r versus their publications, respectively. The color here indicates the number of researchers overlapped in each data point. The darker the color is, the more researchers overlap in the data point. The low Pearson correlation coefficients indicate that D_1 and D_r are both weakly correlated with publication. In this figure, the distribution of D_1, D_r and publication are shown as well (color figure online)

difference between the representative works and regular papers of Nobel Laureates is more different than the scientist with the same h index.

In Fig. 6a, we further show the relation between D_r and h index by plotting the average $\langle D_r \rangle$ of the researchers with the same h index. One can see that $\langle D_r \rangle$ first increases with h and then reaches a plateau around 0.25. Due to this property, if a researcher's D_r is larger than 0.25, he or she can be considered to be above average. In this section, we aim to compare D_r of researchers in different countries. We identify for each country the total number of its researchers (denoted as N) and the fraction of these researchers with D_r larger than a threshold Δ (denoted as p). Δ is set to be 0.25 to ensure that the selected researchers' D_r values are above the overall average. We then plot in Fig. 6b countries' ranks based on N against their ranks based on p. 19 major countries which publish most frequently in APS journals are shown. We can observe that there is a positive relation between the ranking based on N and the ranking based on p. This indicates that for a country with many researchers, not only it will have a high number of researchers with D_r larger than Δ, but also it will have a high fraction of researchers with D_r larger than Δ. Meanwhile, several outliers are identified. Countries likes China and India have a high rank in N yet with a low rank in p, implying these countries have many people working in academia but only small proportion of them have the representative works which have a bigger difference than their regular papers. On the other hand, though Switzerland and Netherland have a small N, many of their researchers have outstanding representative works.

Discussion

Usually when researchers apply for a grant or promotion, they are always requested to list several representative papers. Some scientists also often list several selected publications in their personal websites. Therefore, it is of great significance to have representative works

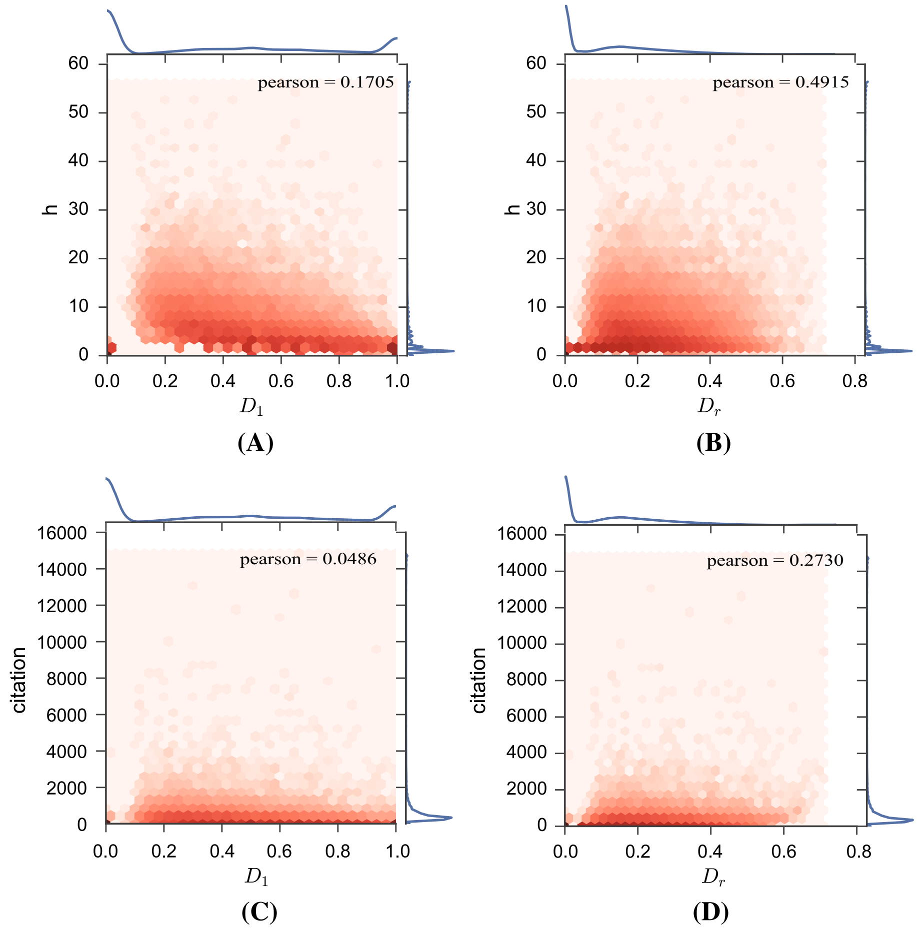

Fig. 4 The scatter plot of researchers' D_1 and D_r versus their h index and total number of citations, respectively. The color here indicates the number of researchers overlapped in each data point. The darker the color is, the more researchers overlap in the data point. The low Pearson correlation coefficients show that D_1 and D_r are both weakly correlated with h index and citation, which indicates that high D_r is not limited in the researchers with large h and many citations. In this figure, the distribution of D_1, D_r, h index and citation are shown as well (Color figure online)

for the development of the personal career and academic reputation. In this paper, we sort the published papers of an author by their citation counts in descending order, which results in a gap between citation counts of any two adjacent papers. The maximum gap is then identified to divide the papers into two groups, with the papers on the left side of this gap regarded as the representative works of the researcher while the papers on the right side of this gap as the regular works. When the maximum gap is close to zero, the representative works have similar impact with the regular works of a scientist. If the maximum gap is large, the impact difference between these two groups of papers are substantial. Based on the maximum gap, we construct a simple indicator D_r to quantify the impact difference between a scientist's representative works and regular works.

Table 3 The basic statistical information of the 28 Nobel laureates in Physics

Id	Year	Authors	No. of papers	No. of citations	h	D_r
1	2013	P. W. Higgs	3	377	3	0.4915
2	2013	F. Englert	13	270	7	0.4917
3	2012	S. Haroche	61	3109	28	0.2426
4	2012	D. J. Wineland	68	3337	29	0.1976
5	2008	Y. Nambu	44	3406	20	0.3564
6	2007	A. Fert	73	2257	21	0.4990
7	2007	P. Grunberg	25	1045	11	0.4579
8	2005	R. J. Glauber	46	3280	25	0.3348
9	2005	J. L. Hall	41	991	17	0.4808
10	2005	T. W. Hansch	100	2016	25	0.4127
11	2004	D. J. Gross	59	3571	25	0.1824
12	2004	F. Wilczek	115	6229	42	0.1692
13	2004	H. D. Politzer	25	1708	16	0.3343
14	2003	A. J. Leggett	61	3758	23	0.2120
15	2002	R. Davis	65	679	14	0.2039
16	2002	M. Koshiba	38	722	12	0.3274
17	2002	R. Giacconi	2	15	2	0.2847
18	2001	W. Ketterle	87	6096	39	0.4940
19	1998	R. B. Laughlin	37	2255	17	0.4687
20	1998	D. C. Tsui	198	5151	39	0.5104
21	1998	H. L. Stormer	66	3017	30	0.5166
22	1997	S. Chu	101	2095	25	0.3013
23	1997	C. Cohen-Tannoudji	24	893	16	0.2663
24	1997	W. D. Phillips	58	1878	26	0.0906
25	1996	D. D. Osheroff	43	760	17	0.2614
26	1996	R. C. Richardson	32	639	17	0.1526
27	1996	D. M. Lee	61	704	16	0.2102
28	1995	M. L. Perl	119	1012	15	0.2968

We apply the selection method and D_r index into the APS dataset. Compared with using the most cited paper of a Nobel laureate to identify his or her representative work, our method can more correctly identify the Nobel prize-winning papers. We also find that D_r index correlates weakly with h index and citation. It means that there is no direct correlation between the individual scientific impact and maximum gap. We also find that the number of representative works selected is related to D_r. When the D_r is small, i.e. D_r is less than 0.15, the large number of selected papers (> 20) can appear. In addition, overall the differences between representative works and nonrepresentative works of Nobel laureates are bigger than their corresponding scientist with similar h index. We calculate the D_r index for each researcher and further at the national level show the statistical results of D_r index.

Based on the citations of a researcher's publications, we only use the maximum gap to select the representative works and ignore some information of other gaps in a histogram of

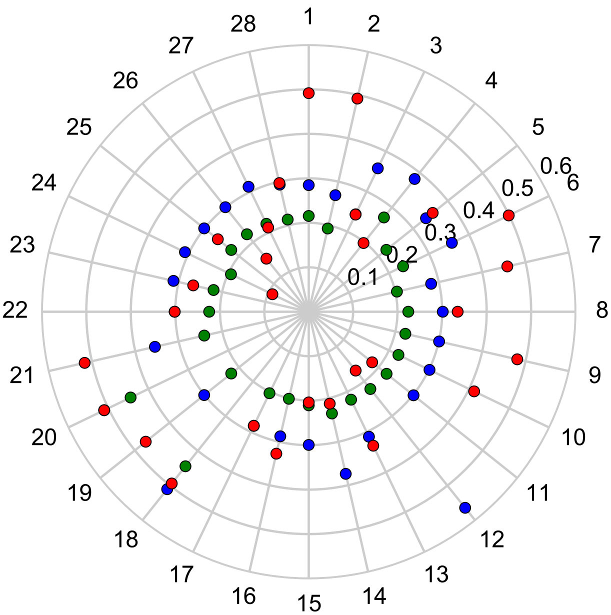

Fig. 5 The D_r index of each Nobel laureate and their corresponding different quantiles of D_r index of the scientists with the same h index. Red dots represent the D_r index of the Nobel Laureates, green dots and blue dots represent the corresponding median and third quartile of D_r index of the scientists with the same h index as the Nobel Laureates, respectively (Color figure online)

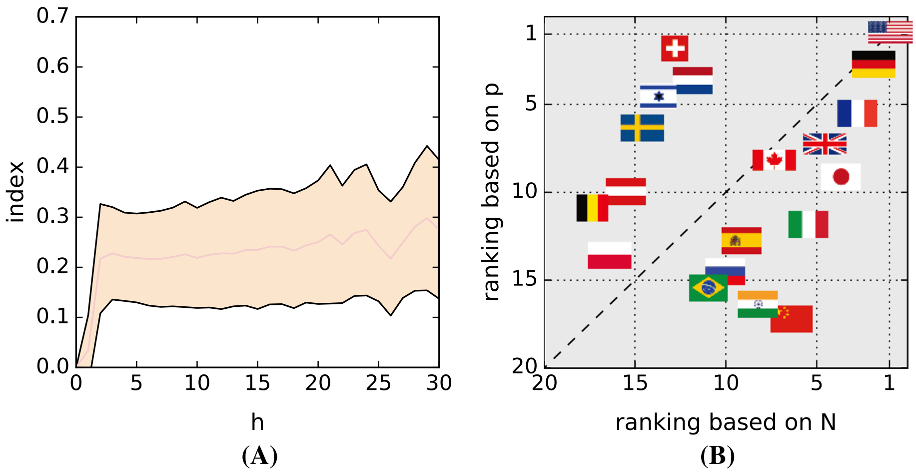

Fig. 6 **a** shows the average $\langle D_r \rangle$ of researchers with the same h index. $\langle D_r \rangle$ first increases with h and then reaches a plateau around 0.25. The shadow represents the standard deviation. **b** shows the scatter plot of countries' ranks based on the total number N of researchers against their ranks based on the fraction p of researchers with D_r larger than a threshold $\Delta = 0.25$. 19 major countries which publish most frequently in APS journals are shown (Color figure online)

sorted papers' citation counts. Therefore, in future we should try our best to propose a more reasonable method to select the representative works of a scientist considering the information of other gaps. The D_r index could quantify the difference between the selected representative works and regular papers. In the analysis, we keep the specific type of publications such as letters and reviews which usually receive more citation and the proposed D_r index might be affected. In addition to physics, our selection method and D_r

index could also be extended to other fields, but the formula of D_r index may need to be adjusted because of different citation patterns existing across scientific fields.

Acknowledgements This work was supported by the National Natural Science Foundation of China (Grant Nos. 61603046 and 61374175) and the Natural Science Foundation of Beijing (Grant No. L160008).

References

Bao, P., & Wang, J. (2018). Identifying your representative work based on credit allocation. *Companion Proceedings of the The Web Conference*, *2018*, 5–6.

Bao, P., & Zhai, C. (2017). Dynamic credit allocation in scientific literature. *Scientometrics*, *112*(1), 595–606.

Bergstrom, C.T., Foster, J. G., & Song, Y. (2016). Why scientists chase big problems: Individual strategy and social optimality. Preprint at arXiv:1605.05822.

Clauset, A., Arbesman, S., & Larremore, D. B. (2015). Systematic inequality and hierarchy in faculty hiring networks. *Science Advances*, *1*(1), e1400005.

Fiala, D., Šubelj, L., Žitnik, S., & Bajec, M. (2015). Do PageRank-based author rankings outperform simple citation counts? *Journal of Informetrics*, *9*(2), 334–348.

Foster, J. G., Rzhetsky, A., & Evans, J. A. (2015). Tradition and innovation in scientists' research strategies. *American Sociological Review*, *80*(5), 875–908.

Garfield, E. (1955). Citation indexes to science: A new dimension in documentation through association of ideas. *Science*, *122*(3159), 108–111.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences of the United States of America*, *102*(46), 16569–16572.

Ioannidis, J., Boyack, K. W., Small, H., Sorensen, A. A., & Klavans, R. (2014). Bibliometrics: Is your most cited work your best? *Nature News*, *514*(7524), 561.

Larsen, P. O., & Von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, *84*(3), 575–603.

Mariani, M. S., Medo, M., & Zhang, Y. C. (2016). Identification of milestone papers through time-balanced network centrality. *Journal of Informetrics*, *10*(4), 1207–1223.

Medo, M., & Cimini, G. (2016). Model-based evaluation of scientific impact indicators. *Physical Review E*, *94*(3), 032312.

Niu, Q., Zhou, J., Zeng, A., Fan, Y., & Di, Z. (2016). Which publication is your representative work? *Journal of Informetrics*, *10*(3), 842–853.

Radicchi, F., Fortunato, S., & Castellano, C. (2008). Universality of citation distributions: Toward an objective measure of scientific impact. *Proceedings of the National academy of Sciences of the United States of America*, *105*(45), 17268–17272.

Schlagberger, E. M., Bornmann, L., & Bauer, J. (2016). At what institutions did Nobel laureates do their prize-winning work? An analysis of biographical information on Nobel laureates from 1994 to 2014. *Scientometrics*, *109*(2), 723–767.

Shen, H. W., & Barabási, A. L. (2014). Collective credit allocation in science. *Proceedings of the National academy of Sciences of the United States of America*, *111*(34), 12325–12330.

Sinatra, R., Wang, D., Deville, P., Song, C., & Barabási, A. L. (2016). Quantifying the evolution of individual scientific impact. *Science*, *354*(6312), aaf5239.

Zeng, A., Shen, Z., Zhou, J., Wu, J., Fan, Y., Wang, Y., et al. (2017). The science of science: From the perspective of complex systems. *Physics Reports*, *714–715*, 1–73.