

# Bibliometric Indices for the Assessment of the Citation Curve Tail

Antonia Gogoglou  
Department of Informatics  
Aristotle University  
54124 Thessaloniki, Greece  
agogoglou@csd.auth.gr

Antonis Sidiropoulos  
Department of Information  
Technology  
Alexander Technological  
Educational Institute of  
Thessaloniki  
57400 Thessaloniki, Greece  
asidirop@it.teithe.gr

Dimitrios Katsaros  
Department of Electrical  
Engineering & Yale Institute for  
Network Science  
Yale University  
USA  
dkatsar@inf.uth.gr

Yannis Manolopoulos  
Department of Informatics  
Aristotle University  
54124 Thessaloniki, Greece  
manolopo@csd.auth.gr

## ABSTRACT

*Scientometrics* is an area recently attracting greater research interest. To a great extent this area has been fertilized by the proposal of the *h-index* (2005), which represents a measure for the quality and quantity of a researcher's impact. The *Perfectionism Index* has been recently proposed aiming at differentiating between 'influentials' and 'mass producers' (2015); the former category produces articles, which are (almost all) with high impact, whereas the latter category produces a lot of articles with moderate or no impact at all. In this paper, we record a number of metrics that are of similar nature, i.e. they shed light into these publishing patterns (influentials vs. mass producers). We carry out a correlation analysis to reveal which metrics are describing the above phenomenon in a similar way and, thus, retain the most descriptive features. Finally, we report the results of an experiment with a dataset consisting of the academic staff of Greek Computer Science/Engineering departments.

## CCS Concepts

•Information systems → Clustering; Digital libraries and archives; Data warehouses; Data analytics;

## Keywords

Scientometrics, *h* index, Perfectionism index, Correlation analysis, PCA analysis

## 1. INTRODUCTION

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from [permissions@acm.org](mailto:permissions@acm.org).

PCI 2015, October 01-03, 2015, Athens, Greece

© 2015 ACM. ISBN 978-1-4503-3551-5/15/10...\$15.00

DOI: <http://dx.doi.org/10.1145/2801948.2801972>

*Scientometrics* is an area recently attracting greater research interest. A seminal paper that founded this area is the one by Eugene Garfield, which defined the notion of *Impact Factor* (IF) to measure the impact of journals [7]. According to Garfield: "In any given year, the impact factor of a journal is the average number of citations received per paper published in that journal during the two preceding years". For example, the IF of JournalX for year 2015 is calculated by the ratio  $A/B$ , where  $A$  is the number of times that all items published in JournalX in 2013 and 2014 were cited by indexed publications during 2015, whereas  $B$  is the total number of "citable items"<sup>1</sup> published by JournalX in 2013 and 2014.

For decades IF was wrongly used as a metric to measure the performance of authors. It was in 2005 that the concept of *h-index* was proposed by Jorge Hirsch as a measure for the quality and quantity of a researcher's impact [11]. In particular, according to Wikipedia: "a scientist has index  $h$  if  $h$  of his/her  $N_p$  papers have at least  $h$  citations each, and the other  $(N_p - h)$  papers have no more than  $h$  citations each".

In the sequel, several variations of the  $h$  index have appeared in the literature in an effort to illuminate a researcher's impact from other angles [1]. Many efforts enhanced the original  $h$  index by taking into account age-related issues [17], multi-authorship [10], fractional citation counting [13], the highly cited articles [5]. Other works explored its predictive capabilities [12], its robustness to self-citations [16], etc. Some of the proposals have been implemented in commercial and free software, such as Matlab<sup>2</sup> and the Publish or Perish software<sup>3</sup>.

These metrics, although informative, they do not take into consideration an author's citation curve as a whole, in particular the curve tail which carries significant information as well [19]. Rosenberg referred to the authors with few cita-

<sup>1</sup>Editorials or letters to the editor are not citable items.

<sup>2</sup><http://www.mathworks.de/matlabcentral/fileexchange/28161-bibliometrics-the-art-of-citations-indices>

<sup>3</sup><http://www.harzing.com/pop.htm>

tions in the upper area and many citations in the tail area as ‘mass producers’, since they have a lot of publications but mostly of low impact [15]. Finally, he referred to the rest of the scientists as the ‘prolific’ ones. Notably, also, the origin of this terminology is quite old (see [4, 6]). In this spirit, the *Perfectionism Index* has been recently proposed aiming at differentiating between ‘influentials’ and ‘mass producers’; the former category produces articles which are (almost all) with high impact, whereas the latter category produces a lot of articles with moderate or no impact at all [18].

In the present work we focus on a set of metrics, that identify weaknesses in the citation curve and thus complement the  $h$ -based metrics to offer a more detailed assessment of publishing patterns and allow for a more delicate distinction among scientists by focusing on previously overlooked behaviors (mass producing). For this set of metrics we carry out a correlation analysis among them and analyze the insight they provide into a scientist’s work as well as the purposes for which they can be used. We also describe how a dataset of the academic staff of Greek computer science and engineering departments has been extracted by using the Microsoft Academic Service. Finally, we report the results of an experiment on this dataset in an effort to reveal researchers which avoid such patters.

## 2. DEFINITION OF TAIL-RELATED SCIENTOMETRIC INDICES

In [18] the Perfectionism Index has been defined (in simplified form) as:

$$PI = h^2 + C_E - C_{TC}$$

where  $C_E$  is the number of citations in the excess  $e$ -area above the core  $h$ -area of size  $h^2$ , whereas  $C_{TC}$  is the complement area of the  $h$ -tail and equals:

$$C_{TC} = h \cdot (P - h) - C_T$$

where  $P$  is the number of publications and  $C_T$  is the number of citations in the curve tail. Figure 1 shows an example of a citation curve and the previously mentioned areas.

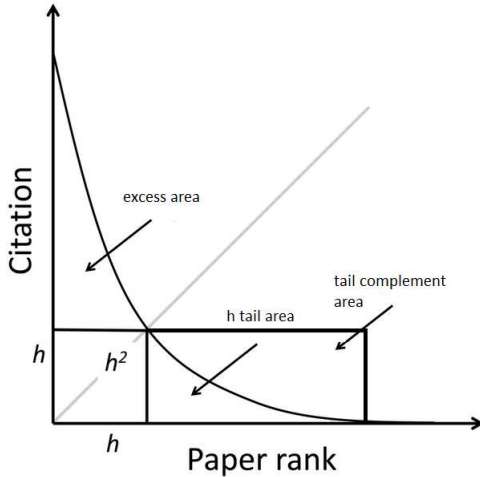


Figure 1: Citation curve and assorted areas

For the experiments conducted in [18] it turned out that the above index takes positive as well as negative values (the

greater the positive value, the more perfectionist an author is; on the contrary, the smaller the negative value, the more mass producer the author is).

### Coefficient $\alpha$ .

Hirsch gives a formula that bridges the number of citations  $C$  with the  $h$  index [11]:

$$C = \alpha \cdot h^2$$

and mentions that usually this  $\alpha$  coefficient takes values in the range [3..5]. Presumably, a researcher with an  $\alpha$  coefficient closer to the value of 3 (5) should be categorized in the class of ‘influentials’ (‘mass producers’, respectively.)

### Distance from maximal $h$ -value.

In an optimal (imaginary) situation the citation curve could be a straight line with slope  $135^\circ$ . It can be easily shown by simple geometry that in this case the above  $\alpha$  coefficient should equal 1, whereas the  $h$  value should be equal to  $\sqrt{C}$ . The difference of the maximal  $h$  value  $\sqrt{C}$  and the actual  $h$  value as a percentage of the maximal  $h$  value allows for comparisons between authors of different scientific impact and age. Thus:

$$\text{normalized distance from maximal } h \text{ value} = \frac{\sqrt{C} - h}{\sqrt{C}}$$

Apparently, the range of values is [0..1]. A researcher with such a value closer to 0 (1) should be categorized in the class of ‘influentials’ (‘mass producers’, respectively).

### Distance from no-length tail.

In another optimal setting, a researcher could have authored  $h$  papers with  $h$  citations each. In this imaginary case, his citation curve should have no  $e$ -area or  $h$ -tail area. To penalize the effect of long tails, we devise the following simple formula, which is again normalized so that researchers of different academic age and productivity level can be compared:

$$\text{normalized distance from no\_length tail} = \frac{P - h}{P}$$

Apparently, the range of values is [0..1]. A researcher with such a value closer to 0 (1) should be categorized in the class of ‘influentials’ (‘mass producers’, respectively).

### Distance from no-area tail.

Consider an identical to the above optimal setting with a researcher who has authored  $h$  papers with  $h$  citations each. A thick tail means that the percentage of citations coming from the tail to the total number of citations is high. In other words, a high percentage of an author’s papers have acquired few or zero citations indicating a non optimal publishing pattern. To penalize the effect of thick tails, we devise the following simple formula, which is again normalized so that researchers of different academic age and impact can be compared:

$$\text{normalized distance from no\_area tail} = \frac{C_t}{C}$$

Again, the range of values is [0..1]. A researcher with such a value closer to 0 (1) should be categorized in the class of ‘influentials’ (‘mass producers’, respectively).

### Coefficient of the power-law.

A quantity obeys a power law if it is drawn from a probability:

$$p(x) \propto x^{-\epsilon}$$

where the exponent  $\epsilon$  is a constant parameter known as scaling parameter and usually lies in the range [2..3].

While several papers have focused on the universality or not of the power exponent of the citation curve [14], our approach is to assign every author with a unique power law coefficient according to the author's individual citation curve. This coefficient represents the level of skewness of the citation curve, meaning that a high power-law exponent indicates a few well cited publications and an abrupt fall in the citation curve leading to a series of low or zero cited papers. This publishing behavior could be considered inconsistent and far from ideal. Thus, we reach an effective and easily interpretable way to characterize an author's publication pattern. It has been well established in previous works that the exponent decreases with an increasing  $h$  value [9]. In other words, the more highly cited the scientist, the lower his/her power-law exponent.

It has been suggested that in practice most empirical phenomena obey a power law only for values of  $x$  higher than  $x_{min}$ . A method for calculating the exponent  $\epsilon$  uses a direct numerical maximization of the logarithm of the likelihood function as follows [3]:

$$L(\epsilon) = -n \ln \zeta(\epsilon, x_{min}) - \epsilon \sum_{i=1}^n \ln x_i$$

To quantify the distance between the estimated distribution and the real data, the Kolmogorov-Smirnov test is used to help optimizing the final result. This process was computerized utilizing the function `plfit` in Matlab<sup>4</sup>.

As can be easily understood to find a realistic exponent a large amount of cited papers would be required; however, a satisfactory estimation can be given for every author as we will present in the next subsection to help identify the level of skewness in the citation curve as an avoidable metric for the author.

The following table summarizes the above list of measures.

Math representation	Description
$PI$	Perfectionism Index
$\alpha$	Hirsch coefficient
$(\sqrt{C} - h)/\sqrt{C}$	Distance from maximal $h$ -value
$(P - h)/P$	Distance from no-length tail
$C_t/C$	Distance from no-area tail
$\epsilon$	Power-law coefficient

**Table 1: Measures associated with the tail of the citation curve**

### 3. DATASET

A dataset of 868 authors members of Greek Universities, who publish in the field of Computer Science/Engineering and Informatics was acquired from the Microsoft Academic Search<sup>5</sup>. From these authors only 345 were included in the final dataset for experimentation, after excluding those with

<sup>4</sup><http://tuvalu.santafe.edu/~aaronc/powerlaws/plfit.m>

<sup>5</sup><http://academic.research.microsoft.com>

less than 2 papers or citations and  $h$  index less than 3 since these authors can be considered as either very young or inactive. The initial and final dataset can be found in the DELAB website<sup>6</sup>. It is noticed that the data from Microsoft Academic Search are not as rich as Google Scholar or Scopus; however, they are publicly available with all the necessary metadata. In particular, for each author the above files consist of the following features:

- Author first name
- Author last name
- Author ID
- Number of citations,  $C$
- Number of papers,  $P$
- Number of citations/paper,  $C/P$
- Number of citations in e-area,  $C_E$
- Number of citations in the tail complement,  $C_{TC}$
- Hirsch  $h$  index
- Hirsch  $\alpha$  coefficient
- Perfectionism Index,  $PI$
- Metric 1,  $(\sqrt{C} - h)/\sqrt{C}$
- Metric 2,  $(P - h)/P$
- Metric 3,  $C_t/C$
- Metric 4, Power-law coefficient  $\epsilon$

### 4. CORRELATION ANALYSIS

Three methods were examined to find out the appropriate correlation analysis method to be used to identify the dependencies between the metrics mentioned above. These methods are: Pearson, Spearman and Kendall correlation [8]. Pearson correlation is widely used in statistics to measure the degree of the relationship between linearly related variables that are normally distributed, whereas Spearman is a non-parametric test apt for measuring the degree of association between two variables without making any assumptions about the distribution of the data. The Kendall correlation is another non parametric coefficient that reverses the ordering of each of the variable vector under question. As shown in Table 2, the distribution of Metric 2 is skewed, whereas Metric 1 and 3 are not as highly skewed; thus, they are closer to the normal distribution. However, the standard deviation is not sufficiently larger than the skewness. Hence, we cannot claim that any of these four metrics is similar to the normal distribution. As a result, the Pearson correlation is further excluded as a metric of correlation analysis. Next, Spearman and Kendall correlations will be examined to identify the relationship between these metrics.

Metrics	Skewness	Standard deviation
Metric 1	-0.089	0.093
Metric 2	-2.484	0.096
Metric 3	0.146	0.170
Metric 4	0.950	0.534

**Table 2: Statistics of the 4 metrics**

A complete correlation analysis based on Spearman's  $\rho$  and Kendall's  $\tau$  was conducted to better display the relationships between the proposed power-law based metric, the other 3 metrics and existing indexes, like  $h$  and  $PI$ , and thus help decipher which metrics add new information to the evaluation of a scientist. A Spearman correlation of  $-0.283$

<sup>6</sup><http://delab.csd.auth.gr>

	$h$	$\alpha$	$PI$	M1	M2	M3	M4	$\rho$
$h$	1	0.097	-0.283	0.097	-0.152	-0.179	-0.134	0.279
$\alpha$	0.097	1	0.003	1	0.284	-0.085	-0.277	0.392
$PI$	-0.283	0.003	1	0.003	-0.730	-0.601	-0.173	0.399
M1	0.097	1	0.003	1	0.284	-0.085	-0.277	0.392
M2	-0.152	0.284	-0.729	0.284	1	0.705	0.148	0.472
M3	-0.179	-0.085	-0.601	-0.085	0.705	1	0.443	0.443
M4	-0.134	-0.277	-0.173	-0.277	0.148	0.443	1	0.350

**Table 3: Correlation matrix based on Spearman method**

	$h$	$\alpha$	$PI$	M1	M2	M3	M4	$\tau$
$h$	1	0.067	-0.221	0.067	-0.108	-0.127	-0.093	0.240
$\alpha$	0.067	1	-0.007	$\sim 1$	0.198	-0.053	-0.195	0.360
$PI$	-0.221	-0.007	1	-0.007	-0.536	-0.417	-0.114	0.329
M1	0.067	$\sim 1$	-0.007	1	0.198	-0.053	-0.195	0.360
M2	-0.108	0.198	-0.536	0.198	1	0.524	0.108	0.382
M3	-0.127	-0.053	-0.417	-0.053	0.524	1	0.327	0.357
M4	-0.093	-0.195	-0.114	-0.195	0.108	0.327	1	0.292

**Table 4: Correlation matrix based on Kendall method**

indicates a negative relationship between the  $h$  index and  $PI$ , mostly due to the negative values of the  $PI$  index for the majority of the authors but a general low correlation without a strict monotonic behavior. The Kendall value is  $-0.221$ , even lower given that it handles the ties in a more efficient way as discussed above. The results of the correlation analysis are complying with the conclusions driven from the distribution function. Tables 3 and 4 contain the Spearman and Kendall correlations respectively for all the discussed metrics, namely, the Hirsch  $h$  index, the Hirsch  $\alpha$  parameter, the  $PI$  index, and the 4 metrics described above, denoted as M1, M2, M3 and M4.

These tables indicate the relationship between the above mentioned 7 metrics and the following conclusions can be drawn:

- The Hirsch  $\alpha$  parameter is absolutely correlated with Metric 1, since they both express a ratio of the author's citations in the  $h$  core to the total number of received citations. The reason why Metric 1 was introduced is that it can be easily normalized and thus can be compared with the other metrics. As can be seen, Metric 1 is the least correlated with all the metrics, with the exception of the  $\alpha$  parameter that is essentially similar. The last columns of the tables indicate the mean correlation of each metric with all others in terms of absolute values. The  $h$  index and Metric 1 seem to be displaying an average correlation of 0.279 and 0.392 respectively. This is expected since Metric 1,  $h$  index and  $\alpha$  parameter are the two measures that do not contain any information about the tail of the citation curve.
- The  $PI$  index shows a high degree of correlation with Metrics 2 and 3, which is rather expected as they are all indexes carrying information about the curve tail. The negative monotonic correlation is justified by the sign of the  $PI$  index; in other words, the higher the value of the metrics, the higher the penalty for the academic performance of the author. This is expressed by the sign of the  $PI$  index.
- Metrics 2 and 3 display moderate to high correlation

with each other and the  $PI$  index. However, they display low correlation with the  $h$  index, which should be expected as they convey a completely different kind of information about the citation curve. Metric 4 is also weakly correlated with the  $h$  index but moderately correlated with the thickness of the tail (Metric 3) which proves the link of the power law coefficient with the tail of the citation curve.

- The higher the mean Spearman correlation  $\rho$ , the more representative are the measures. The meaning is that they contain information of other indexes as well and thus they give a complete picture of academic performance. In this case, the  $PI$  index and Metrics 2 and 3 can be considered representative of the citation curve, with the  $h$  index being the least correlated to all the other measures. In other words, the  $h$  index seems to ignore the information the metrics are expressing and focuses on a different part of the curve.
- The same can be reported for the Kendall correlation but with smaller values of correlation.

The correlation matrix (with either correlation technique) of Table 3 after removing the  $\alpha$  parameter was subjected to PCA analysis. The correlation values between  $h$ ,  $PI$  and the four metrics were found to be statistically significant based on the  $p$  value of  $p < 0.05$ .

## 5. PRINCIPAL COMPONENT ANALYSIS

The above correlation matrix was subjected to Principal Component Analysis, which by means of eigenvalue decomposition identified 5 orthogonal principal components. PCA is considered an appropriate methodology to identify the information conveyed by bibliographic metrics and the clusters that are formulated [2]. The resulting principal components ranked according to the degree by which they explained the variances in the correlation matrix of the 6 metrics are listed in Table 5.

It can be seen that the first 3 components yield an explanation of the 84.62% of the total variance, meaning that the data can be fully expressed in these first 3 components. Since the original variables were only 6, a 3D representation

	PC1	PC2	PC3	PC4	PC5
Percentage of variance explained	42.57	23.81	18.24	10.63	4.14
Cumulative percentage	42.02	66.38	84.62	95.25	99.39

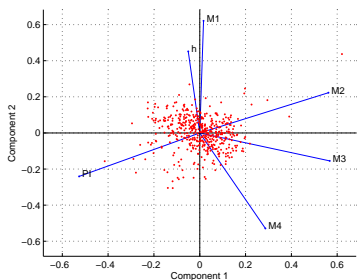
**Table 5: Variance explained - Components in latent order**

of the initial correlation matrix would be appropriate. However, usually in the case of many variables that need to be expressed, the first 2 principal components are considered sufficient to explain the variance in the data, thus, leading to a 2D representation. In this case, 66.38% of the variance would be explained by the first 2 components. The original correlation matrix of 6 variables and their correlations is now transformed to a new coordinate system based on the 2 or 3 first principal components, thus making it easier to visualize the indexes and what common features they share.

Table 6 shows the principal component coordinates of the 6 indexes based on 2 Principal Components, whereas Figures 2 and 3 illustrate the principal component coordinate space and the 6 variables in 2D and 3D plots, respectively. The red dots are the original scores of the authors according to the 6 indexes transformed in the principal component coordinates. The blue lines depict the initial variables, i.e. the metrics.

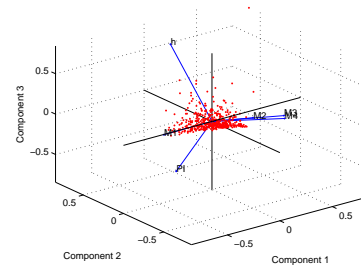
	PC1	PC2	PC3
<i>h</i>	-0.050	0.451	0.767
<i>PI</i>	-0.526	-0.240	-0.323
<i>M1</i>	0.016	0.620	-0.454
<i>M2</i>	0.561	0.225	-0.254
<i>M3</i>	0.567	-0.155	-0.056
<i>M4</i>	0.286	-0.528	0.177

**Table 6: Principal components of the 6 metrics**



**Figure 2: 2D Principal Components**

From the above figures and tables we conclude that the 1st principal component displays positive coefficients with the 4 metrics. This is the reason why they are directed to the right half of the plot. The largest coefficient though is with metrics 2 and 3, indicating that PC1 offers information about publishing policies that negatively affect academic influence. Metrics 2, 3 and 4 can be seen as one cluster that contains the same quality of information meaning the characteristics of power law distribution and the tail of the citation curve. On the other hand, the 2nd component displays strong positive coefficients with *h* index and Metric 1 (in other words,



**Figure 3: 3D Principal Components**

the Hirsch  $\alpha$  parameter). It would be safe to assume that PC2 focuses on information about the *h* core and attributes of the citation curve that positively affect academic performance. These two metrics are another cluster and the final cluster is *PI* on its own with a negative coefficient with both components due to its penalizing attitude and the rankings it assigns to our set of authors, which are vastly different from rankings based on the other metrics.

Should this analysis be expanded to a larger number of features, more clusters could appear and through this kind of visualization groups of indexes could be identified allowing for a better and more accurate feature selection among bibliographic indexes. Depending on the task at hand, either the most representative index, in other words the most centered and correlated one, or a set of indexes identified from different groups based on specific needs could be chosen to evaluate to the fullest the academic work of a researcher.

## 6. SPECIFIC REMARKS ON THE DATASET

From the described dataset in Section 3, we extracted the top-20 researchers according to their *h* value. We ranked them according to Metrics 1 to 4 and *PI*. We can easily detect that there is a significant variation in the resulting ranks especially between the Metric 1 and the other ones, which is in compliance with the correlation analysis described before. In the computation of the rank based on the 4 metrics it is clear that if we handle the ties by placing the authors with the same score in the same rank two conclusions can be drawn. First of all, Metric 3 offers better distributed rankings, with fewer ties, meaning that it is more precise in identifying smaller differences between the authors. On the other hand, Metric 1 appears to be more concentrated with a considerable amount of authors receiving the same score. Metric 4 shows little variation among scientists, which should be expected since the differences in the exponent of the power law distribution are small but of great value to help distinguish between authors that have similar values in the other indexes.

Table 7 shows the specific sample of top authors, where the data values represent rankings in each specific metric. It can be easily understood that there are significant differences in the rankings when using these four metrics, which concurs to the original correlation analysis. Consequently, the use of all metrics is justified since they convey different information aspects about the evaluation of the scientific work of an author. Even between Metrics 2 and 3 that were found correlated, there is a certain variance in the ranking output. This clearly demonstrates that the distribution of the two

metrics is quite similar; however, the absolute score values of each metric can deviate significantly indicating that the two metrics convey different aspects of the same information, i.e. the tail of the citation curve.

	<i>h</i>	<i>PI</i>	M1	M2	M3	M4
I. Pitas	44	20	5	19	20	17
L. Tassioulas	41	15	18	15	5	8
<b>D. Gunopoulos</b>	39	2	9	5	11	13
Y. Ioannidis	39	10	3	6	9	19
<b>M. Garofalakis</b>	37	3	2	2	8	18
T. Sellis	36	17	12	13	12	14
<b>C. Courcoubetis</b>	30	1	20	7	2	2
M. Vazirgiannis	26	7	17	11	11	3
Y. Manolopoulos	26	19	16	20	19	8
G. Polyzos	26	16	13	17	7	3
Y. Theodoridis	26	12	10	12	3	12
N. Mamoulis	26	13	1	8	18	20
S. Christodoulakis	25	11	8	10	13	6
M. Koubarakis	24	6	7	4	14	15
V. Christophides	23	5	15	3	6	10
E. Pitoura	22	14	11	14	16	9
P. Maragos	21	18	6	18	17	4
I. Tollis	21	8	19	16	4	5
<b>Y. Kotidis</b>	21	4	14	1	1	1
P. Vassiliadis	21	9	4	9	15	16

**Table 7: Top-20 researchers according to *h* index and ranking according to position based on metrics M1-M4**

Authors with bold characters in Table 7 seem to outperform with respect to these bibliometric indices. For example, Yannis Kotidis is ranked 1st according to 3 criteria, Costas Courcoubetis is ranked 1st in 1 criterion and 2nd in another 2 criteria, whereas Minos Garofalakis is ranked 3rd in 1 criterion and 2nd in other 2 criteria. In addition, Dimitrios Gunopoulos shows a balanced presence in all metrics. Thus, these researchers have a publishing behavior that categorizes them in the class of influential scientists.

## 7. CONCLUSIONS

Recently the Perfectionism Index has been proposed as a means to categorize researchers in two classes: influential vs. mass producers. Along the same line, in this paper we have provided 4 new simple criteria to assess weak characteristics of publishing patterns. We have also extracted a dataset of Greek researchers from university departments in Computer Science/Engineering and Informatics, which is publicly available. Finally, we have performed a correlation analysis to keep the most important from the independence point of view. These criteria augment the long list of the necessary features to assess the performance of researchers. Lastly, we extract 3 researchers out of the specific dataset, which can be characterized as influential scientists according to the above criteria.

## 8. REFERENCES

- [1] S. Alonso, F. J. Cabrerizo, E. Herrera-viedma, and F. Herrera. h-Index: A review focused in its variants, computation and standardization for different scientific fields. *Journal of Informetrics*, 3:273–289, 2009.
- [2] J. Bollen, H. V. de Sompel, A. Hagberg, and R. Chute. A principal component analysis of 39 scientific impact measures. *PLoS ONE*, 4:e6022, 2009.
- [3] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *SIAM Rev.*, 51(4):661–703, Nov. 2009.
- [4] S. Cole and J. R. Cole. Scientific Output and Recognition: A Study in the Operation of the Reward System in Science. *American Sociological Review*, 32(3):377–390, 1967.
- [5] L. Egghe. Theory and practise of the *g*-index. *Scientometrics*, 69(1):131–152, 2006.
- [6] G. J. Feist. Quantity, quality, and depth of research as influences on scientific eminence: Is quantity most important? *Creativity Research Journal*, 10(4):325–335, 1997.
- [7] E. Garfield. Citation indexes for science; a new dimension in documentation through association of ideas. *Science*, 122(3159):108–111, July 1955.
- [8] G. D. Garson. *Correlation, 2012 Edition*. Statistical Associates Publishers, Blue Book series, Asheboro, NC, 2012.
- [9] P. R. Gupta H., Campanha J. Power-law distributions for the citation index of scientific publications and scientists. *Brazilian Journal of Physics*, 35:981–986, 2005.
- [10] J. Hirsch. An index to quantify an individual’s scientific research output that takes into account the effect of multiple coauthorship. *Scientometrics*, 85(3):741–754, 2010.
- [11] J. E. Hirsch. 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, 2005.
- [12] J. E. Hirsch. Does the *h* index have predictive power? *Proceedings of the National Academy of Sciences*, 104(49):19193–19198, 2007.
- [13] D. Katsaros, L. Akritidis, and P. Bozanis. The *f* index: Quantifying the impact of coterminal citations on scientists’ ranking. *Journal of the American Society for Information Science and Technology*, 60(5):1051–1056, 2009.
- [14] G. J. Peterson, S. PressÃ, and K. A. Dill. Nonuniversal power law scaling in the probability distribution of scientific citations. *Proceedings of the National Academy of Sciences*, 107(37):16023–16027, 2010.
- [15] M. Rosenberg. A biologist’s guide to impact factors. *Arizona State University, Technical Report*, <https://dx.doi.org/10.7287/peerj.preprints.477v1>, 2011.
- [16] M. Schreiber. Self-citation corrections for the hirsch index. *EPL (Europhysics Letters)*, 78(3), 2007.
- [17] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos. Generalized hirsch *h*-index for disclosing latent facts in citation networks. *Scientometrics*, 72(2):253–280, 2007.
- [18] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos. Ranking and identifying influential scientists versus mass producers by the perfectionism index. *Scientometrics*, 103(1):1–31, 2015.
- [19] F. Y. Ye and R. Rousseau. Probing the *h*-core: an investigation of the tail-core ratio for rank distributions. *Scientometrics*, 84(2):431–439, 2010.