Evaluating Greek Departments of Computer Science/Engineering using Bibliometric Indices

Dimitrios Katsaros[1] Vassilios Matsoukas[2] Yannis Manolopoulos[2]

[1]Department of Computer & Communication Eng
University of Thessaly
Volos, GREECE
dkatsar@inf.uth.gr

[2]Department of Informatics
Aristotle University
Thessaloniki, GREECE
manolopo@csd.auth.gr

Abstract

The evaluation of the scientific work though scientometric indicators has long attracted significant scientific interest, but recently has become of ground practical and scientific importance, due to its use in faculty promotions, funding allocation, and so on. The proposal of the h-index for individuals by Jorge Hirsch in 2005 has taken the world of research assessment by storm, and subsequently various flavors of it have been proposed to cure its age-ignorant behavior, to extent its application in institutional evaluation, etc. This article analyzes the performance of the Greek Departments of Computer Science/Engineering using a series of bibliometric indicators based on the h-index, and using also some traditional indicators that measure productivity. The purpose of the analysis is to provide a new dimension toward a fair assessment of merit for better conception of their image in the Greek society and public opinion and the Greek scientific community as well.

Keywords: Informetrics, Scientometrics, Impact Factor, h-index, University rankings, Research evaluation.

1. Introduction

The academic community is working since the 1970s with the aim of evaluating scientific work in an unbiased and fair way. Having defined metrics for this purpose, we can use them for faculty recruitment or promotion, for prize awarding, for funding allocation, for comparison of personal scientific merit, etc. Similarly, the estimation of a publication forum's (journal or conference) quality is of particular interest, since it impacts the scientists' decisions about where to publish their work, the researchers' preference in seeking for important articles, and so on.

Currently, a specific academic community, i.e. the International Society for Scientometrics and Informetrics [ISSI, 2008], has been formed maintaining journals and conferences, which deal with Qualitative Studies on research topics related to the present study, and, in general, with Mathematical, Statistical, and Computational Modeling and Analysis of Information Processes.

However, except of the academic interest, these issues attract enormous commercial interest as well. To start ab ovo, the Institute for Scientific Information (ISI) was founded by Eugene Garfield (1960). It was acquired by Thomson Scientific & Healthcare (1992), became known as Thomson ISI and now as Thomson Scientific [Thomson, 2008]. It is a component of the multi-billion dollar Thomson Reuters Corporation. The ISI specialty is citation indexing and analysis, a field pioneered by Garfield who introduced the concept of Impact Factor for journals [Garfield, 1972]. The ISI maintains citation databases covering thousands of academic journals available via the Web of Knowledge database service [Web of Knowledge, 2008]. This database allows a researcher to identify which articles have been cited most frequently, and who has cited them. The ISI also publishes a list of highly cited researchers, one of the factors included in the Academic Ranking of World Universities published by Shanghai Jiao Tong University [Wikipedia, 2008].

The ISI publishes annual Journal Citation Reports, which list an impact factor for each tracked journal. Notable, due to the huge development of digital libraries, the research efforts have been intensified lately, and, thus, we have witnessed a blossom of this field. However, despite the theoretical and practical popularity of the Impact Factor, within the scientific community, the Impact Factors play a large but controversial role in determining the kudos attached to a scientist's published research record. Some reasons for this attitude are [Sidiropoulos and Manolopoulos, 2005]:

- Each scientific field is divided in certain static areas, which do not reflect the scientific evolution and, particularly, the dramatic evolution of computer science.
- In each area, only a set of journals is selected for journal evaluation. Thus, the representative value of the selected journals is questionable.
- Although any such set is dynamic and updated periodically, this update is done in a subjective way (not to say commercial),.
- In some cases, irrelevant journals (e.g. technical vs. popular) are grouped in a certain area leading to erroneous results.
- Conferences, books and technical reports are considered. Recently, Thomson planned to expand its services with conference evaluation.
- It is not for free neither for libraries nor for individuals, and lastly
- For a certain period until our days, the Impact Factor has been the only quantitative measure in the field.

Recently, a new metric has been proposed in the literature to evaluate an individual's scientific contribution, the h-index, named after Hirsch [Hirsch, 2005]. In very short time, h-index attracted the interest of the scientific community for several theoretical and practical issues. Is this metric just another yet overestimated metric or a useful tool for objective evaluation? Several variations have been proposed to alter some deficiencies of the metric or to better enlighten some specific instances of a researcher's contribution. Nowadays, of particular interest is the work by Ann-Wil Harzing, who has set a popular website with free downloadable software to asses an individual's scientific work with a number of h-index variations [Harzing, 2008]. Recently also, the metric of h-index has been generalized to assess the scientific contribution of departments, universities or programs of studies. Among others, we note the recent works of [Arencibia et al. (2007), Da Luz et al. (2008), Egghe (2008), Egghe and Rao (2008), Ma et al. (2008), Ruane and Tol (2008), Schubert (2007)].

Since one of the topics of the current dispute in Greek universities is the concept of evaluation [Lazaridis, 2007], we contribute to this debate by trying to evaluate the Greek Departments of Informatics/Computer Science (CS) with 4 years of studies and the Departments of Computer Engineering (CE) with 5 years of studies. Having recognized the complications of this task and following the suggestions of [van Raan, 2005] we use advanced metrics, i.e., the notion of h-index, apart from the simple metrics that are solely based on productivity. The rest is organized as follows. Next we define the h-index and all other concepts to be used in the sequel, whereas Section 3 describes the performed experimental work. Section 4 draws some conclusions regarding the specific aim of the paper, i.e. the evaluation of Greek Departments of CS/CE. Furthermore, we bring into light some fallacies regarding the reputation of these departments in the Greek public opinion. Section 5 concludes the paper.

2. Definitions

In this section we define the concepts that have been mentioned in the previous and are going to be used in the sequel. The interested reader can refer to [Harzing, 2008] for more aggregated information.

The Impact Factor is the average number of citations in a year, given to those articles in a journal that were published during the two preceding years. For example, the 2008 impact factor of a journal is A/B, where A is the number of times articles published in 2006-7 were cited in indexed journals during 2008, and B is the number of articles published in 2006-7 [Wikipedia, 2008].

Hirsch's h-index aims at providing a robust single-number metric of an academic's impact, combining quality with quantity. For example, a researcher has h-index equal to h if h of his/her N_p articles have received at least h citations each, and the rest (N_p-h) articles have received no more than h citations [Hirsh. 2005].

Egghe's g-index aims at improving the h-index by giving more weight to highly-cited articles. In particular, a researcher has g-index equal to g if g of his most cited articles (taken descendingly from his/her N_p articles) have received collectively at least g^2 citations [Egghe 2006].

The Contemporary h-index (hc-index) aims at improving the h-index by giving more weight to recent articles, thus rewarding academics who maintain a steady level of activity. For technicalities see [Sidiropoulos at al. 2007].

The Individual h-index (hi-index) divides the standard h-index by the average number of authors in the articles that contribute to the h-index, to reduce the effects of co-authorship [Batista et al. 2006]. Another variation of the Individual h-index instead of dividing the total h-index, it first normalizes the number of citations for each article by dividing the number of citations by the number of authors for that article, and then calculates the h-index of the normalized citation counts. This approach, called hi-norm, is much more fine-grained than that of [Batista et al. 2006] and, thus, it is believed that it accounts more accurately for any co-authorship effects that might be present and that it is a better approximation of the per-author impact, which is what the original h-index set out to provide [Harzing, 2008].

According to [Bihui 2007], the Age-weighted Citation Rate (AWCR) measures the average number of citations for the articles which contribute to the h-index, adjusted for the age of each individual article. In particular, the count of citations for a specific article is divided by the age of the article. AR-index is the square root of AWCR for the articles which contribute to the h-index. In [Harzing, 2008] a variation of AR-index is proposed, the AW-index, according to which AWCR measures the average number of citations to the entire body of a researcher's work, and not only for the articles which contrivute to the h-index. Finally, AWCRpA (=per author AWCR) performs a normalization not with respect to the age of the article but with respect to the number of authors per article.

3. Experimental setting

Evaluation of the Greek Departments of Computer Science/Engineering (in short, DCS/DCE) using bibliographic indices relies on the evaluation of the academic staff of each respective department. The latter task can be safely performed by using the tool Publish of Perish, downloadable from [Harzing, 2008]. This tool retrieves and analyzes academic citations by using [Google Scholar, 2008] to obtain the raw citations, and then to analyze them and present the following statistics:

- Total number of papers and total number of citations
- Average number of citations per year, per paper and per author
- Average number of papers per author and authors per paper
- Hirsch's h-index and related parameters

- Egghe's g-index
- Sidiropoulos's contemporary h-index
- The age-weighted citation rate
- Two variations of individual h-indices.

Having calculated the above metrics for all the individuals, then in a similar way we can calculate the same metrics at a departmental level. In particular, the h-index of a specific department is equal to *h* if *h* of his/her N_p faculty members have a value of *h* as h-index, and the rest (N_p-h) faculty members have no more than *h* value as h-index. Similar is the definition of a department's g-index, hc-index, hi-index and hi-norm. However, to calculate a department's AW-index we take the square root of the sum of the AWCR of the faculty members of the specific department.

In our survey we included 17 departments hosting 552 staff members of all ranks (i.e., Profs, Associate Profs, Assistant Profs and Lecturers). We excluded recently founded departments since they do not have an adequate number of staff members, so that reliable statistics can be extracted. Due to often errors when translating Greek names to English ones, it was first necessary to avoid name ambiguities. In practice, each person's name in English was double-checked though his personal webpage. In several cases, there were no personal webpages, a fact that absorbed significant energy for resolution. To locate an individual's publications record, his/her last name was used in the field "Authors" of Publish or Perish, whereas his/her initials were used in the "The phrase" field to guarantee that his/hers (and only his/hers) works could be retrieved. Also, the retrieval was performed with the selection "Years of Publication between 1960 and 2007", because in several instances publications dated since 200 years were retrieved due to errors in the accessed databases.

Subsequently, these data were categorized according to the institution of each individual. Table 1 shows the statistics for the staff of the Department of Informatics of the Aristotle University of Informatics. In the last column, symbols A, B, C and D stand for the 4 professorship ranks. In the last line of the same table we depict the aggregated statistics for this department. Then, by collecting the aggregated statistics for all departments we produced a table, which is omitted for brevity. Instead, here we report Table 3 where each column shows each department's order in relation to the whole set of 17 ones. More specifically, the 4th and 5th columns depict the ranking of each department from the productivity point of view (i.e., number of authored papers and number of citations received). The subsequent 6 columns are related to the qualitative bibliographic metrics as described in the previous section and generated by the Publish or Perish tool. The last but one column (i.e. points) gives the order of the popularity of each department according to prospective students' preferences. It is noted that this figure corresponds to the minimum number of points that a high-school graduate has to get at national-level exams to be accepted by each department (data downloaded from [Ministry of Education, 2008]). Finally, the last column depicts the order of each department in terms of size (i.e. total number of staff members), where the largest department gets rank 1. It is noted that these data (acquired from Web during January 2008) are highly dynamic both in terms of the individuals involved as well as in terms of their contributions. However, due to the exponential nature of h-index and its variations we reasonably anticipate that a few changes in the meanwhile can not alter the validity of our results. On the other hand, the quality of our data is based on the quality of the publically available data maintained by departments and individuals. Finally, we note that the records of Table 2 are sorted according to the h-index value, whereas in case of a tie, we sort according to the g-index value and then according to the hc-index value.

4. Comments

By searching for regularities and irregularities in Table 2, we can draw interesting observations. In the sequel, we will use the university names in short instead of the department names. As expected, the National Technical University of Athens, the University of Athens and the University of Patras stand high in the list according to their academic performance. However, from the latter table it comes up that the University of Crete is ranked 1^{st} in (almost) all examined bibliometric indices, even though it is not the most productive (ranked 8^{th}) due to its medium size (ranked 11^{th}). Interestingly, also, this institution is not very popular among the prospective students (ranked 14^{th}), which might be attributed to its distance from the Greek mainland.

When comparing the values of the first two numeric values, we could expect that the corresponding values per each department should be about the same. However, we observe deviations in both ends. For example, we remark that the University of Crete and the Athens University of Economics & Business show an outstanding relative quality (in terms of the number of citations) in comparison to their production (in terms of articles authored). In the opposite end, we classify the University of Patras (Department of Electrical & Computer Eng) and the University of Thrace.

A closer examination at the hc-index reveals another interesting fact. The Athens University of Economics & Business, the University of the Aegean and the University of Piraeus demonstrate a considerable deviation between the h-index value and the hc-index value. This practically means that these institutions show a significant academic performance during the very recent years, maybe due to a very successful recent recruitment policy.

When examining hi-index values in comparison to the h-index values, we remark significant deviations in the case of the University of Patras (Department of Computer

Eng & Informatics) on one hand, and the cases of the Athens University of Economics & Business and the Aristotle University (Department of Electrical & Computer Eng) on the other hand. The first case stems from the fact that the author list of the articles authored by the respective faculty members is rather long, whereas the contrary holds for the latter two cases.

Another anticipated pattern is that the larger size of an institution (i.e. number of faculty members), the higher its expected production (i.e. articles authored). As in the previous, this pattern has fortunate and unfortunate exceptions. The most striking fortunate exceptions to this pattern are the University of Thessaly. On the contrary, at he other end we meet the University of Thrace and the University of Macedonia. The reader can seek for a couple of other fortunate and unfortunate exceptions as well.

It is theoretically well-known that 5-year faculties (engineering departments) are more popular that 4-year ones (science departments). This is practically verified in the table. However, this ranking of public preferences (i.e. by prospective students and their families) is far from agreeing with the ranking according to the academic competence. For example, except the University of Crete, also the Aristotle University (Department of Informatics) is under-preferred. On the contrary, the University of Thrace, the University of Macedonia and the Aristotle University (Department of Electrical & Computer Eng) are over-preferred although their academic performance in terms of bibliometric indices is not that high as their popularity.

5. Conclusion

Ranking of research institutions by bibliometric methods is a tool for research evaluation that steadily gains popularity among policy makers, public media and scientific world. Although there are significant problems with this methodology (e.g., data collection and cleansing), it is generally perceived as a valuable tool in measuring academic excellence. This article takes a first step in providing a ranking of the Greek Departments of Computer Science/Engineering using advanced bibliometric indicators to assess their performance. The investigation was able to mark the “best” department, to record some generic patterns and most importantly to discover “outliers” to these patterns.

References

Arencibia-Jorge R., Barrios-Almaguer I., Fernández-Hernández S. and Carvajal-Espino R. (2007). *Applying successive h-indices in the institutional evaluation: a case study*. **Journal of the American Society for Information Science and Technology**, Vol.59, No.1, pp.155-157.

Batista P., Campiteli M., Kinouchi S. and Martinez A. (2006). *Is it possible to compare researchers with different scientific interests?* **Scientometrics**, Vol.68, No.1, pp.179-189.

Bihui J. (2007). *The AR-index: complementing the h-index*. **ISSI Newsletter**, Vol.3, No.1, p.6.

Da Luz M.P., Marques-Portella C., Mendlowicz M., Gleiser S., Coutinho E.S. and Figueira I. (2008). *Institutional h-index: the performance of a new metric in the evaluation of Brazilian Psychiatric Post-graduation Programs*. **Scientometrics**, to appear.

Egghe L. (2006). *Theory and practice of the g-index*. **Scientometrics**, Vol.69, No.1, pp.131-152.

Egghe L. (2008). *Modelling successive h-indices*. **Scientometrics**, to appear.

Egghe L. and Rao R. (2008). *Study of different h-indices for groups of authors*. **Journal of the American Society for Information Science and Technology**, Vol.59, No.8, pp.1276-1281.

Garfield E. (1972). *Citation analysis as a tool in journal evaluation*. **Science**, Vol.178, pp.471-479.

Google Scholar (2008). http://scholar.google.gr/

Harzing (2008). http://www.harzing.com

Hirsch J.E. (2005). *An index to quantify an individual's scientific research output*. **Proceedings National Academy of Sciences**, Vol.102, No.46, pp.16569–16572.

ISSI (2008). http://www.issi-society.info.

Lazaridis T. (2007). http://www.sci.ccny.cuny.edu/~themis/greekuniversityreform

Ma R., Ni C. and Qiu J. (2008). *Scientific research competitiveness of world universities in computer science*, **Scientometrics**, Vol. 76, No. 2, pp. 245-260, 2008.

Ministry of Education, 2008. http://www.ypepth.gr/

Anthony F.J. van Raan (2005). *Fatal attraction: Conceptual and methodological problems in the ranking of universities by bibliometric methods*. **Scientometrics**, Vol. 62, No. 1, pp.133-143.

Ruane F. and Tol R.J. (2008). *Rational (successive) h-indices: an application to economics in the Republic of Ireland*. **Scientometrics**, Vol.75, No.2, pp.395-405.

Schubert A. (2007). *Successive h-indices*. **Scientometrics**, Vol.70, No.1, pp.201-205.

Sidiropoulos A. and Y. Manolopoulos (2005). *A new perspective to automatically rank scientific conferences using digital libraries*. **Information Processing & Management**, Vol.41, No.2, pp.289–312.

Sidiropoulos A., Katsaros D., and Manolopoulos Y. (2007). *Generalized h-index for Disclosing Latent Facts in Citation Networks*. **Scientometrics**, Vol.72, No.2, pp.253-280.

Thomson (2008). http://scientific.thomson.com

Web of Knowledge (2008). http://www.webofknowledge.com

Wikipedia (2008). http://en.wikipedia.org/wiki/Institute_for_Scientific_Information

Table 1. *Data for the academic staff of the Department of Informatics of Aristotle University*

A/A	Name	Papers	Citations	Cites/ year	Cites/ paper	Cites/ author	Papers/ author	Authors/ paper	h-index	g-index	hc-index	hI-index	hI-norm	AW-index	Prof rank
1	I PITAS	681	6623	264,92	9,73	3371,69	295,74	2,63	39	70	25	17,89	26	25,34	A
2	Y MANOLOPOULOS	304	2393	108,77	7,87	881,18	122,85	2,80	20	44	15	7,02	14	16,83	A
3	M HATALIS	133	577	19,90	4,34	204,36	46,39	3,23	14	21	8	4,17	7	7,83	A
4	I VLAHAVAS	187	660	31,43	3,53	259,46	71,35	2,91	12	18	10	4,24	7	10,04	A
5	A POMPORTSIS	149	475	36,54	3,19	172,25	50,13	3,06	10	18	8	3,45	7	8,8	A
6	C KARANIKAS	41	91	3,25	2,22	46,83	21,89	2,12	5	7	3	2,27	4	2,93	A
7	I TSOUKALAS	76	110	3,24	1,45	35,35	27,41	3,24	4	9	4	1,07	3	3,94	A
8	GI PAPADIMITRIOU	150	560	31,11	3,73	275,17	59,66	2,87	12	20	8	6,26	8	8,98	B
9	H KARATZA	117	228	10,36	1,95	171,50	89,24	1,46	8	9	6	5,33	6	6,29	B
10	P LINARDIS	28	44	1,47	1,57	15,98	9,95	2,96	4	5	3	1,45	2	1,97	B
11	C KOTROPOULOS	195	1000	52,63	5,13	346,99	69,53	3,06	16	27	11	4,83	9	10,88	C
12	I STAMELOS	87	365	16,59	4,20	120,55	26,66	3,31	11	17	9	3,18	6	8,27	C
13	N BASSILIADES	98	374	23,38	3,82	152,90	39,42	2,87	11	16	9	4,03	6	8,41	C
14	L ANGELIS	61	292	12,70	4,79	100,68	20,97	3,11	8	15	8	2,37	5	7,38	C
15	A VAKALI	115	288	16,94	2,50	147,02	52,66	2,85	8	13	7	3,56	5	7,18	C
16	N NIKOLAIDIS	133	1022	56,78	7,68	394,74	49,38	3,22	13	30	8	3,60	6	10,57	D
17	A NANOPOULOS	76	465	22,14	6,12	157,04	24,44	3,29	10	20	9	3,33	8	8,29	D
18	N PLEROS	56	312	39,00	5,57	69,00	12,70	4,71	10	16	9	2,04	4	7,42	D
19	A PAPADOPOULOS	53	265	22,08	5,00	114,66	19,77	3,02	8	15	5	3,05	5	5,68	D
20	T TSIATSOS	72	197	19,70	2,74	63,90	22,01	3,44	8	10	6	2,21	4	5,84	D
21	G TSOUMAKAS	31	94	11,75	3,03	36,34	11,98	3,03	6	7	6	2,40	3	4,52	D
22	N LASKARIS	34	80	3,48	2,35	25,51	10,19	3,59	5	7	4	1,47	2	3,37	D
23	D VRAKAS	32	74	7,40	2,31	27,30	11,81	2,78	5	7	4	1,67	3	3,11	D
24	P NICOPOLITIDIS	39	124	17,71	3,18	73,25	11,70	3,59	4	10	3	1,60	3	4,9	D
25	P KATSAROS	33	40	1,18	1,21	19,03	16,49	2,52	4	4	3	1,78	2	1,86	D
26	A MILIOU	15	52	3,47	3,47	18,17	5,08	3,07	3	7	2	1,00	2	2,26	D
27	D POLITIS	24	19	1,46	0,79	7,16	11,18	2,83	3	4	2	1,13	2	1,81	D
28	S DIMITRIADIS	27	35	1,59	1,30	13,66	12,22	2,93	3	4	2	0,90	2	1,79	D
29	N ATREAS	14	27	2,70	1,93	11,17	6,83	2,36	3	4	2	1,13	2	1,84	D
	TOTALS	**3061**	**16886**	**29,09**	**3,68**	**252,86**	**42,40**	**3,00**	**10**	**14**	**8**	**3,44**	**7**	**45,41**	

Table 2. *Final data set for the Greek Departments of Computer Science/Engineering*

	DEPARTMENT	UNIVERSITY	Papers	Cites	h-index	g-index	hc-index	hi-index	hi-norm	AW-index	Points	# of Profs
1	Computer Science	U Crete	8	2	1	1	1	1	1	2	14	11
2	Electrical & Computer Eng	Tech U Athens	1	1	1	1	2	2	2	1	1	1
3	Informatics & Telecommunications	U Athens	3	3	3	3	3	3	3	4	4	5
4	Computer Eng & Informatics	U Patras	4	6	4	4	3	9	3	6	5	7
5	Electrical & Computer Eng	U Patras	2	7	4	4	6	6	6	7	3	2
6	Informatics	U Thessaloniki	5	5	6	6	6	7	6	3	11	7
7	Informatics	Athens U Economics	10	4	7	7	3	4	3	5	8	7
8	Electrical & Computer Eng	U Thessaloniki	6	8	7	7	6	5	6	8	2	3
9	Electronics & Computer Eng	Tech U Crete	9	9	7	7	9	8	6	10	10	14
10	Informatics	U Ioannina	11	11	10	10	9	11	11	11	13	14
11	Electrical & Computer Eng	U Thrace	7	12	10	11	9	10	6	12	6	3
12	Information & Communications Eng	U Aegean	14	14	12	11	9	14	13	14	15	10
13	Computer & Communications Eng	U Thessaly	12	10	12	11	14	13	11	9	9	17
14	Informatics	U Piraeus	13	13	12	14	9	12	13	13	12	12
15	Applied Informatics	U Macedonia	15	15	15	15	14	15	15	15	7	6
16	Informatics	TEI Athens	16	17	16	16	16	17	17	17	16	12
17	Informatics	TEI Thessaloniki	17	16	16	17	16	16	15	16	17	16