Research in Progress Paper presented at the 13th Conference of the International Society for Scientometrics and Informetrics (ISSI). 4 - 7 July, 2011. Durban, South Africa.
http://www.issi2011.uzulu.ac.za/

Investigating the nature of scientific reputation

Cristhian Parra[1], Fabio Casati[1], Florian Daniel[1], Maurizio Marchese[1], Luca Cernuzzi[2], Marlon Dumas[3], Peep Kungas[3], Luciano García-Bañuelos[3], Karina Kisselite[3]

[1] *{parra, casati, marchese, daniel}@disi.unitn.it*
Dipartimento di Ingegneria e Scienza dell'Informazione, University of Trento
Via Sommarive, 14 38123, Trento, Italy

[2] *lcernuzz@uca.edu.py*
Universidad Católica "Nuestra Señora de la Asunción" – Departamento de Electrónica e Informática
Tte. Cantalupi y Tte. Villalón. Barrio Santa Ana. Asunción – Paraguay

[3] *{marlon.dumas, peep.kungas, luciano.garcia}@ut.ee, karina.kisselite@gmail.com*
Institute of Computer Science, University of Tartu, Estonia
J Liivi 2. Tartu 50409, Estonia

Abstract
Excellence or quality are often regarded as the holy grail of science, and it is the main goal driving scientist to pursue research that will influence the direction of their fields. This excellence, however, has no standard definition and varies across disciplines, and even from person to person, making it difficult to evaluate research and reputation of researchers. In this work, we introduce the general problem of studying the nature of reputation in the context of computer science, providing preliminary results on its relation with bibliometric indicators and hints for future experiments that will foster a better understanding of reputation in the scientific domain.

I. Introduction
The underlying mechanisms that rule researchers mind when they evaluate their peers are, in a way, hidden to most of us and can be regarded as a mix of objective (e.g. bibliometrics) and subjective (e.g. affiliation) criteria. Understanding how this two dimensions are combined to form a general rule of evaluation, that result in a measure of reputation**,** is an open issue that, if better understood, could foster the creation of better models for research assessment.
Currently, the most cost-effective and thereby most widely used methods to support evaluation of research are based on bibliometric indicators (e.g. h-index (Hirsch 2005), g-index (Egghe 2006) among others). The common intuition is that these indicators are very good summarizers of the impact of a scientist.
However, recent studies have indicated that bibliometric indicators alone are not enough to fully assess the quality of the work of scientists. Criteria for measuring quality have different weights and meaning across communities (Lamont 2009) among other issues (Bollen 2009, Adler 2008, Priem, Taraborelli et. al. 2010, Brody and Harnad 2005, Laloë 2009, Priem and Hemminger 2010, Bollen 2007).
In this article we present a methodology to study how scientific reputation develops in the head of evaluators and report the initial experimental results on the correlation between bibliometric indicators and perceived reputation of computer scientists.

II. Evaluation in Science
Research in this area has been focused on quantitative methods, developing indexes to measure scientific impact, mainly based on citation analysis. Bibliometric indexes, mainly based on citation and publication analysis, are the emerging and most used tools to perform research evaluation of people and their contributions, making the citation the cornerstone of scientific impact (Garfield and Welljams-Dorof 2004). The *h-index* (h number of publications with h or more citations) deserves a special mention in this field for being the one that has had a greater reach across communities for its simplicity to summarize both productivity and

impact in one single number. The literature on indexes is extense, but most of them are derived from h-index's initial insight (Bar-Ilan 2008). On the other hand, one of the most important methods for running an assessment process consists on forming panels of experts (e.g. Peer Review) that will evaluate the subjects of the assessment (e.g. people, contributions, etc.). These panels usually rely again on quantitative methods as the first approach for evaluation. However, very often, panellists will use also a number of more subjective and qualitative methods, i.e. interviews, paper reviews etc.

Alternative metrics (such as number of downloads, number of views, etc.), based neither on citations nor publication counts, are also being developed to capture other type of interactions (e.g. social bookmarking) (Priem, Taraborelli et. al. 2010, Priem and Hemminger 2010, Bollen 2007 and *The MESUR project*). These new metrics are the answer to problems of only relying in citations for research assessment (Adler 2008).

As for *Journals,* the most renowned metric is the Journal Impact Factor (JIF) based on ISI Thomson Scientific Database and defined as the average number of citation per article in the journal over a two-year period (Nazri and Halif 2007). The "Eigenfactor" (Bergstrom 2007) has been proposed as a better metric for Journals, estimating of percentage of time that library users spend with a journal. This estimate is computed by modeling a random walk in the citation graph following an algorithm similar to PageRank (Brin and Page 1998).

As for Conferences and Institutions, evaluation is mainly based on rankings coming from trustworthy institutions that run evaluation processes. Some examples of these institutions are the *Research Assessment Exercise (RAE)* in UK and the *Excellence in Research for Australia (ERA) Initiative*. Panels of experts compile these rankings defining their own criteria for the evaluation and using a mix of objective and subjective criteria (e.g. esteem indicators of RAE). In this matter, the Wikipedia (see *"College and university rankings"*) offers one of the most comprehensive summaries of rankings for institutions.

III. Collecting Reputation Opinions

The ultimate goal of our ongoing work is to understand the nature of reputation in Computer Science. To reach this goal, we need first to study how different types of information, characterizing researchers and their work, might be related to the notion of reputation based on the opinion of peers. Understanding the nature of reputation in academia, with the focus on the Computer Science, poses two sets of orthogonal challenges. The first related to the (i) collection of data and the second related to the (ii) reverse engineering of the reputation logic.

The first set of challenges has to do with harvesting the right information that would allow us to explore what are the features (objective or subjective criteria) that actually influence researchers when they are assessing their peers. Objective criteria include information that can be measured and analyzed using quantitative methods (e.g. number of publications, h-index or number of awards and grants). Subjective criteria include information whose evaluation measure lies on the "eye of the beholder" (e.g. affiliation, nationality, fields of interest, recommenders). Most of the information we need is readily available on the web, but not in the format which could be easily processed by programs. Bibliometrics, for example can be calculated based on data we get from sources like Google Scholar, DBLP, CiteSeer, Web of Science and many others. Subjective criteria are more challenging because first we need to define what information we could qualify in this category and then we need to search sources to get it. The source for this information could be researcher's homepages, conferences or university sites, etc.

Furthermore, we will also need already available reputation opinions. This information is usually not readily available, or might be implicit in some sources (e.g. the result of recruitment competitions, peer review data). The best we can do here is directly asking

researchers about what is their own opinion about other researchers, and then try to extrapolate from this the reputation that could be used as an input for future estimations.
The second set of challenges has to do with using the information we have to derive and represent what are the features and algorithms that best estimate researchers' reputation in a particular domain. This is, reverse engineering the reputation logic that researchers have in their heads.
In this paper we focus on the first set of challenges and we provide some preliminary results plus hints for solution towards understanding reputation in the Computer Science domain.

3.1. Experiments

To study whether there is or not a relation between bibliometric indicators and perceived reputation, we needed (1) to find sources of reputation information for a set of researchers and then (2) to compute bibliometric indicators for the same set of people. In order to obtain the reputation information, we followed two different approaches:

1. A survey asking about research impact and deployed in several conferences of Computer Science.
2. Crawling results from research position contests in Italy (MIUR[1]) and France (CNRS[2]), which are produced by selection committee that runs a voting procedure between candidates.

To obtain the second type of information, i.e. bibliometric indicators for the people involved in the survey and candidates evaluated in the research contests, we implemented our own script to compute bibliometric indicators using results from Google Scholar searches and ReaderMeter.
Once all the information have been collected, correlation analysis was performed using Kendall-tau method to compare rankings resulting from reputation ratings and rankings resulting from bibliometric indicators. In the case of Italian research contests, it was not possible to run a correlation analysis due to the fact that reputation rankings obtained from this source were only pairs of one selected candidate and one candidate put on a waiting list. For this reason, for the Italian dataset we computed the percentage of success that each metric had in predicting the first place of the contest.

3.2. Survey

A scientific reputation survey was designed within the EU project LiquidPub[3] and deployed in several conferences over a set of candidates relevant for that conference.
Each survey consisted on a sample of 40 candidates taken from Jens Palsberg's top h-index researchers list[4]. Half of the sample was computed according to a measure of closeness to the target conference. This closeness was based on the distance within co-authorship networks of evaluated researchers with respect to others that published in the same conference.
In total, 8 surveys were implemented and deployed in conferences such as *BPM (Business Process Management), ICWE (Web Engineering)* and *VLDB (Very Large Databases)*, getting a total of 77 answers in a period of 3 months of being online[5]

---

[1] http://reclutamento.murst.it/sessioni_2008/scrutini_precedenti.php
[2] http://intersection.dsi.cnrs.fr/intersection/resultats-cc-en.do
[3] http://liquidpub.org/
[4] http://www.cs.ucla.edu/~palsberg/h- number.html
[5] http://reseval.org/survey

In each survey, data gathered included some information such as *age*, *position* and *gender* of the voter, plus his or her answer to the request of rating the scientific impact and relevance of the research output of each candidate.

3.3. Competitions Data

The second approach for getting reputation information consisted of getting the results of contests for research position for Italy and France.

In the case of Italy, available data at MIUR site was from 2008 and included, for each contest, the pair of selected candidates where one was the winner of the contests and the other was the second place. For 208 contests pairs where both candidates had at least one recorded citation (from a total of 333 pairs), we later calculated which percentage of the times bibliometric indicators succeeded on predicting the first place of the contest.

In the case of France, CNRS data included a list of more than 1000 researchers participating in different contests whose result were published in the form of a ranking of 2 or more people.

In both cases, our complete analysis has been hindered from the lack of the list of researchers that were not included in the final list of selected candidates. This is, in other words, the lack of information about the real losers of the contests. In both cases, however, we are trying to obtain the complete dataset to improve our analysis.

IV. Results

4.1. Results of Surveys

Analysis of reputation ratings in the survey compared to bibliometric indicators showed, for all conferences, a stable pattern of correlation coefficients below the threshold for considering that there is correlation between the variables in study.

For a correlation to be considered true, the correlation coefficient has to be greater than 0.5 (positive correlation) or less than -0.5 (negative correlation). Figure 1 shows the value of these coefficients for each metric we have calculated, based on the aggregated results from all the surveys. The maximum correlation is that of the h-index available in Palsberg's own website, but is still only 0.33.

Although we do not show all the charts and results in this paper, the same pattern is present in all individual surveyed conferences with the highest correlation equal to 0.60 (the only one higher than 0.5), which was found in the BPM conference with the metric *"Number of Publications (from DBLP)"*.

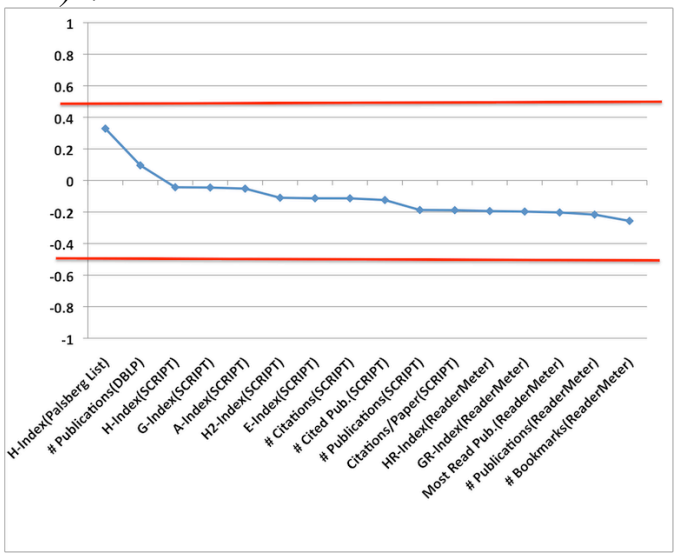

Figure 1. Correlation coefficients between reputation and bibliometric indicators from Surveys

4.2. Results of Recruitment process in CNRS

Correlation analysis of CNRS rankings shows the same pattern of independence (i.e. very low correlation) we encountered in the surveys. As with the previous, all coefficients are near to zero. This being said, we still need to extend this dataset with the names of researchers that were eliminated on early phases of the selection process of CNRS, organized in three stages in order to complete our analysis.

The main problem we faced in this analysis is that very often, researchers in this dataset had few records available in Google Scholar making hard to calculate their bibliometric indicators.

4.3. Results of Recruitment process in Italy

In the case of the Italian research contests, Table 1 shows percentage of cases in which the winner has a lower indicator than the second place (W < S), the winner has a better indicator than the second place (W > S) and finally where indicators are the same (W = S).

Table 1: Bibliometric Indicators performance to forecast Italian Contest results

	H-Index	Citation Count	Cited Publications
W < S	47.1% (98)	56.2% (117)	50.5% (105)
W > S	38.9% (81)	39.4% (98)	47.6% (99)
W = S	13.9% (29)	4.33% (9)	1.92% (4)

We report here only those indicators that had the better performance, which are the h-index, the total citation count and the number of cited publications. As the table shows, no indicator have a performance better than 50%, which means that they predict the result of the contest the same as a random selection. These results, however, need to be validated by further extending the Italian dataset to cover researchers that were eliminated earlier in the process of selection.

V. Discussion

The general intuition is that researchers with high number of publications and high indicators of relevance of their research reach highest scientific reputation. Nevertheless, based on our preliminary results, the general conclusion is that bibliometric indicators and reputation have little or none correlation. This being said, more experiments and analysis need to be done in order to better explain the reasons of this phenomena and whether it affects to all domains in the same manner. These results serve as motivation to continue with a line of work that would eventually allow us to reverse engineering the reputation logic in academia.

5.1. Future Work

These preliminary results of our experiments serve as motivation for future work on understanding how reputation is built in the realm of academia. In the first place, there are two basic types of data we need to collect: data to derive features information (e.g. publication and citation records, affiliation, awards, etc.) and data to derive reputation information (e.g. ratings in a survey, opinions of peers, peer reviews results, etc.). Regarding these challenges, our approach will be focused on (1) designing more surveys about reputation and scientific impact; (2) crawling of contests for research positions, social interactions of researchers in social networks, scientific data and metadata, etc. that could serve to compute relevant features of researchers; and (3) leveraging the power of the crowd in order to get information that is hard or cannot be obtained in an automated manner (i.e. using a similar approach than that of the mechanical turk, see https://www.mturk.com/mturk/welcome).

The second set of challenges has to do with the understanding how this information is combined to built reputation. Our approach will be focused on (1) running statistical analysis over collected data looking for correlation between reputation and features; (2) apply data mining techniques to get descriptive and predictive models of how features influence reputation; and (3) apply social network analysis techniques to understand how data on features and reputation varies across communities.
This work marks the starting point of a long line of research to better understand reputation in science. We believe that this research is necessary to foster better models of evaluation in research that will eventually lead to better and fairer methods to measure scientific impact.

Acknowledgments

This work has been supported by the EU ICT project LiquidPublication, under FET-Open grant number 213360.

References

Adler, R., Ewing J. (Chair) & Taylor, P. (2008). Citation statistics. *Citeseer.* Vol. 58, June 2008 AD. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.73.6511&rep=rep1&type=pdf.

Bar-Ilan, J. (2008). Informetrics at the beginning of the 21st century – A review. *Journal of Informetrics,* 2(1):1–52.

Bergstrom, C. T. (2007). Eigenfactor: Measuring the value and prestige of scholarly journals. *C&RL News*, 68(5):314-316.

Bollen, J., Rodriguez, M. & Van de Sompel. H. (2007). MESUR: usage-based metrics of scholarly impact. *Proceedings of the 7th ACM/IEEE-CS joint conference on Digital libraries,* pp. 474. ACM.

Bollen, J., Van de Sompel, H., Hagberg, A. & Chute, R. (2009). A principal component analysis of 39 scientific impact measures. *PloS one,* 4(6):e6022.

Brin, S. & Page. L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer networks and ISDN systems*, 30(1-7):107-117, 1998.

Brody, T. & Harnad, S. (2005*).* Earlier web usage statisticsas predictors of later citation impact. *CoRR,* abs/cs/0503020.

Egghe, L. (2006). An improvement of the h-index: the g-index. *ISSI Newsletter* 2, no. 1: 8-9. http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.93.4627&rep=rep1&type=pdf.

Garfield, E. & Welljams-Dorof, A. (2004). Of Nobel class: A citation perspective on high impact research authors. *Theoretical Medicine and Bioethics,* 13(2):117–135, 1992.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *Proceedings of the National Academy.* 102, no. 46: 16569-16572. http://www.pnas.org/cgi/content/full/102/46/16569?gca=102%2F46%2F16569&sendit=Get+All+Checked+Abstract(s)&amp.

Laloë, F., & Mosseri, R. (2009). Bibliometric evaluation of individual researchers: not even right... not even wrong!. *Europhysics News* 40, no. 5: 26-29. doi:10.1051/epn/2009704. http://www.europhysicsnews.org/10.1051/epn/2009704.

Lamont, M. (2009). How professor's think. Cambridge, MA: *Harvard University Press.*

Imran M., Marchese, M., Ragone, A., Birukou, A., Casati, F. & Laconich, J. (2010). ResEval: An Open and Resource-oriented Research Impact Evaluation tool.

Nazri, M. & Halif, A. (2007). Journal Impact Factor.

Priem, P. G. J., Taraborelli, D. & Neylon, C. (2010). Alt-metrics: A manifesto.

Priem, P. G. J. & Hemminger, B. H. (2010). Scientometrics 2.0: New metrics of scholarly impact on the social Web. *First Monday,* Volume 15, Number 7 - 5 July 2010. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/viewArticle/2874/257