# All for One or One for All? Analyzing Collaboration Patterns in Research Environments

Mario Cataldi, Luigi Di Caro, and Claudio Schifanella

Université Paris 8, Paris, France University of Torino, Torino, Italy {m.cataldi@iut.univ-paris8.fr, dicaro@di.unito.it, schi@di.unito.it}

Abstract. Researchers, their scientific publications and their research projects are often object of evaluation, from different points of view, for many different purposes. However, even if different metrics have been proposed in literature, they usually assume the co-authorship to be a proportional collaboration between the researchers, missing out their relationships and their change on time along the career. In this work, we propose an application that makes use of a novel metric for evaluating and comparing researchers by taking into account the co-operations among them and estimate their reciprocal dependence degree along time. This application can help comparing and ranking researchers based on his/her demonstrated independence, along his/her whole career, with respect to the surrounding research community.

Keywords: Bibliometrics, Collaboration Patterns, Authors Ranking

## 1 Introduction

Bibliometric indicators are increasingly used to evaluate scientific careers based on personal publication records. The simple number of papers published by an author rather than the received citations are still common ways to capture both the quantity and the impact of an author's set of works. However, these methods do not capture the actual contribution of a researcher within a research network. In this respect, it has been much discussed whether co-authors should have all the same value in quantifying the impact of a paper. In [7], for example, the author first pointed out the problem of *undeserved coauthorship*. In [4] it has been stated that further efforts have to be done in this direction. However, the simple analysis of the position of an author in the list is not enough [5]. Indeed, this generalizes over something that is actually unknown. Which are the rules governing the position of a person in the authors list? An objective and universally-recognized point of view on that simply does not exist.

In light of this, the pure information about the publication records of a researcher often results insufficient for a fair evaluation of scientific profiles because they do not take into account many factors, as the relationships between the authors and their relative scientific influences, which should be directly considered in the evaluation process.

In fact, especially when these measures are used for recruitment purposes, it is highly relevant to analyze the scientific dependencies among authors in order to estimate the capacity of an author to work and produce research outcomes *without* the people that assisted his or her work until that time.

A research collaboration can be indeed defined as a two-way process where individuals and/or organizations share learning, ideas and experiences to produce together scientific outcomes. Collaborations are necessary because of the evident difficulty for individual scientists to conduct several groundbreaking research on their own. For this, one of the key aspect of a successful researcher is the development of a large, active, network of collaborators that can help the researcher to bring new solutions and propose, continuously, novel ideas and approaches to the research community. On the other hand, evaluation of individuals needs a sort of inverse process with the primary goal of understanding the role of each researcher, and his/her specific impact on the research community, in this collaborative environment.

In light of this, following the work presented in [2], the main goal of this study is to introduce a novel indicator for measuring the dependence among scientists by analyzing their co-authorship network and their shared outputs.

With these goals in mind, based on the entire DBLP bibliographic database, in this paper we present a web platform (available at *http://d-index.di.unito.it*), which allows the user to study the scientific profile of each researcher and analyze, through several dynamic visualization tools, the evolution of the impact of each collaboration on his/her scientific output.

# 2 Formalization of Scientific Collaborations in Publication Networks

Based on the previous theoretical works proposed in [1,3], in this paper, we make use of a formalization of the *co-authorship network* that represents the environment in which a researcher has produced his/her scientific outcomes.

Given two collaborating researchers (also called authors along the paper),  $r_i$ ,  $r_j$  and their common scientific network  $N_{r_i,r_j}^t$ , defined as the set of researchers who collaborated with them, the autonomy of their collaboration  $a_{r_i,r_j}^t$  at time t is calculated as:

$$a_{r_i,r_j}^t = \begin{cases} 0 & \text{if } N_{r_i,r_j}^t = \emptyset \\ \frac{1}{\sum_{r_k \in N_{r_i,r_j}^t} \sum_{x=1}^{c(r_k,O_{r_i,r_j}^t)} \frac{1}{x}} & \text{if } N_{r_i,r_j}^t \neq \emptyset \end{cases}$$

where the function  $c(r_k, O_{r_i, r_j}^t)$  returns the number of times a researcher  $r_k$  coauthored a paper with both  $r_i$  and  $r_j$  at time t. The higher the autonomy the more independent the work of  $r_i$  and  $r_j$  is from their research environment. We then define the *dependence value* of  $r_i$  on the collaboration with  $r_j$  as  $d_{r_i \to r_j}^t$  as

$$d_{r_i \to r_j}^t = \frac{p_{r_i, r_j}^t}{p_{r_i}^t} \times \frac{a_{r_i, r_j, N_{r_i}}^t + a_{r_j, \neg r_i, N_{r_i}}^t}{a_{r_i, r_j, N_{r_i}^t}^t + a_{r_j, \neg r_i, N_{r_i}}^t + a_{r_i, \neg r_j, N_{r_i}}^t}$$

where p is a productivity score (number of published works) of a is the autonomy score. The dependence value  $d_{r_j \to r_i}^t$  ranges from  $\theta$  to 1; in particular,  $d_{r_i \to r_j}^t \approx 0$  indicates that the dependence of  $r_i$  on  $r_j$ , at the time t, is negligible, while a  $d_{r_i \to r_j}^t \approx 1$  highlights the contrary.

Thus, given the complete set of dependence values, for each year and relative to each co-author, we calculate the researcher's *dependence trajectory*, by calculating the standard deviation, along the time, of each dependence value, for each co-author, from the optimal attended value of 0 (which would mean a dependence score of 0; i.e., the production of the considered researcher is independent from the collaboration with the considered co-author). In a sense, we aim at evaluating the overall independence of a researcher from the surrounding community. More formally, given a researcher  $r_i$ , we define his/her dependence trajectory  $\overrightarrow{d_{r_i}} = \{sd_{r_i}^t, sd_{r_i}^{t+1}, \cdots, sd_{r_i}^{t+n}\}$ , where  $sd_{r_i}^t$  is calculated as

$$sd_{r_i}^t = \sqrt{\frac{\sum_{r_k \in N_{r_i}} (d_{r_i \to r_k}^t)^2}{|N_{r_i}|}}.$$

We can use these values to properly compare, and rank researchers with similar characteristics. More in detail, we provide a radar chart that can rank the independence performance of a considered researcher with respect to those who have i) similar career length, ii) similar number of publications, iii) similar number of co-authors.

#### 3 Web Application and Real Case Scenario

In this section, we introduce our application for analyzing, comparing and ranking scientific collaboration patterns of researchers. The web application is available at http://d-index.di.unito.it. As data input, we considered the DBLP data set<sup>1</sup>.

The proposed application permits to search for any author indexed by DBLP and to take a preview, through several features and visualizations, of his/her scientific profile and her/his collaboration history over time. The user can analyze the evolution over time of each scientific collaboration for a searched researcher. It is also possible to The system can visualize the evolution of the dependence of a researcher on the support of each co-author along the career. With this chart, it is also possible to select/deselect additional co-authors to make further analyses and comparisons. The application also provides a dynamic visualization chart

<sup>&</sup>lt;sup>1</sup> http://dblp.uni-trier.de/db



**Fig. 1.** A screen-shot, taken from *http://d-index.di.unito.it*, that shows the depedence trajectory and the ranking of the very well known Dr. Faloutsos against similar authors.

(called "time-lapse") which allows the user to focus on a specific time interval and/or a subset of co-authors.

Finally, the proposed tool tool also allows to compare and rank the overall independence of an author, along his/her whole career, with the whole research community. Please also notice that it is also possible to compare the considered researcher against others (even if they do not share the same time career). This visualization permits to focus on how much the entire production of a researcher can be considered dependent on the interactions with her/his local community.

The presented demo can be used to analyze each researcher in the entire DBLP community by also considering similar profiles (with parameters such as number of papers, number of co-authors, and length of career).

### References

- L. D. Caro, M. Cataldi, and C. Schifanella. The d-index: Discovering dependences among scientific collaborators from their bibliographic data records. *Scientometrics*, 93(3):583–607, 2012.
- M. Cataldi, L. D. Caro, and C. Schifanella. Ranking researchers through collaboration pattern analysis. In Machine Learning and Knowledge Discovery in Databases -European Conference, ECML PKDD 2016, Riva del Garda, Italy, September 19-23, 2016, Proceedings, Part III, pages 50-54, 2016.
- L. Di Caro, M. Cataldi, M. Lamolle, and C. Schifanella. It is not what but who you know: A time-sensitive collaboration impact measure of researchers in surrounding communities. In *Proceedings of the 24th International Conference on World Wide Web*, WWW '15 Companion, pages 995–1000, New York, NY, USA, 2015. ACM.
- 4. R. Hunt. Trying an authorship index. Nature, 352(6332):187-187, 1991.
- J. Imperial and A. Rodríguez-Navarro. Usefulness of Hirsch's h-index to evaluate scientific research in Spain. *Scientometrics*, 71(2):271–282, 2007.
- R. M. Slone. Coauthors' contributions to major papers published in the ajr: frequency of undeserved coauthorship. AJR. American journal of roentgenology, 167(3):571–579, 1996.