Ranking Researchers through Collaboration Pattern Analysis

Mario Cataldi, Luigi Di Caro, and Claudio Schifanella

Université Paris 8, Paris, France
University of Torino, Torino, Italy
RAI Research Centre, Torino, Italy
{m.cataldi@iut.univ-paris8.fr, dicaro@di.unito.it,
claudio.schifanella@rai.it}

Abstract. The academic world utterly relies on the concept of scientific collaboration. As in every collaborative network, however, the production of research articles follows hidden co-authoring principles as well as temporal dynamics which generate latent and complex collaboration patterns. In this paper, we present an online advanced tool for real-time rankings of computer scientists under these perspectives.

Keywords: Bibliometrics, Collaboration Patterns, Authors Ranking

1   Introduction

Scientists are object of evaluation for funding allocation and career promotions. The discovery of leading scientists is an important task that simple statistics over long publication records may miss. In this demo, we present an online tool for analyzing researchers under a *collaborative* perspective by studying and ranking their capacity to maintain the same quality/quantity levels in different research environments.

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 a scientist's work. However, these measures represent only an evaluation of what is knowable from simple database searches. Still, these numbers actually make strong assumptions on the co-authorship of the research works in terms of how proportional the collaboration was among the co-authors. In a sense, scientists may look favorably good if working in a dynamic and active research environment. On the contrary, they may result unfairly below par due to modest research collaborators.

In literature, a number of related concepts have been presented, such as the *undeserved co-authorship* [4] and the *scientific relevance* [1]. Instead, our proposed application system is oriented to the study of what *collaboration* means. A research collaboration can be defined as a two-way process where individuals and/or organizations share learning, ideas and experiences to produce joint scientific outcomes. Collaborations are intrinsically necessary to the production of

complete and groundbreaking research. In the light of this, one of the key aspect (and more demanded in recruitment scenarios) of a successful researcher is the development of a large and active network of collaborators that helps researchers bring new solutions within the research community.

The presented online tool is able to automatically compare scientists by deeply analyzing their local co-authorship networks and how they have been crucial in the production of research articles over time. Along the paper, we will present both the theoretical and algorithmic parts of the tool as well as the set of available features which can be freely tested on `http://d-index.di.unito.it`.

2 Background: Formalization of Scientific Collaborations

Based on the theoretical works proposed in [2] and [3], for this demo application, 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), 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:

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

where the function $c(r_k, O_{r_i,r_j}^t)$ returns the number of times a researcher r_k co-authored 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

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

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 *0* 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

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

In words, the system detects anomalies in the collaboration patterns with respect to the attended behavior. Researchers, in fact, are expected to increment their collaboration network over time becoming independent from their single collaborations.
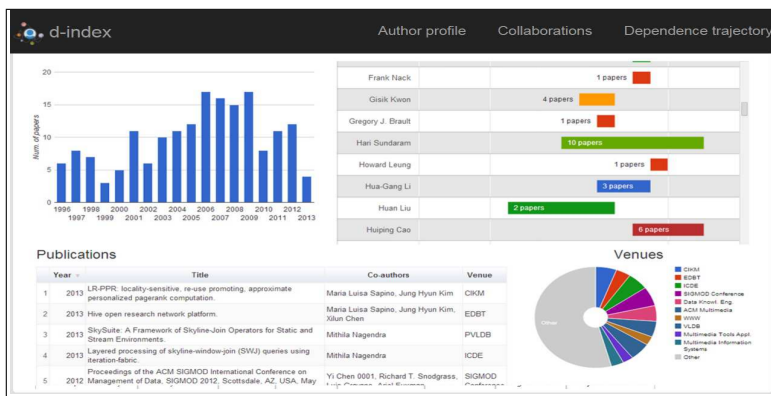

Fig. 1. Profile of a researcher in the presented application demo.

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. We also provide a comparison with the active researchers and the whole community. Finally, we will integrate a feature to compare researchers with respect to topics automatically extracted from publication titles.

3   Application and Demo Scenario

In this section, we present our application, available at *http://d-index.di.unito.it*, for analyzing, comparing and ranking scientific collaboration patterns of researchers. As data input, we considered the DBLP data set[1], containing information about 1,717,211 authors and 3,268,812 scientific papers[2].

As shown in Figure 1, the proposed application initially allows to search for any author indexed by DBLP and to analyze his/her scientific profile and her/his collaboration history over time (through several features and visualizations). Then, the online demo provides the following analyses:

---

[1] http://dblp.uni-trier.de/db
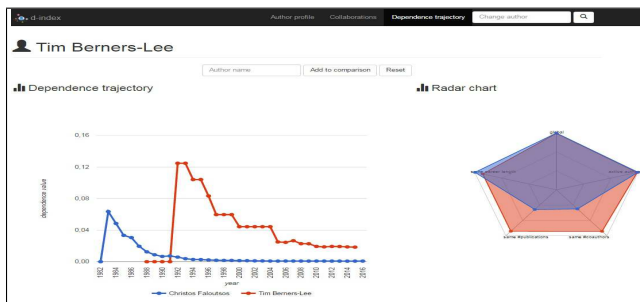[2] Information updated at May 2016.

Fig. 2. Another screen-shot, taken from *http://d-index.di.unito.it*, that compares collaboration patters of Sir Tim Berners-Lee and Dr. Christos Faloutsos.

- **Collaborations over time.** The user can analyze the evolution over time of each scientific collaboration for a searched researcher.
- **Collaboration Pattern Analysis.** The system can visualize the above-mentioned dependence patterns through a curve metaphor, mapping 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.
- **Temporal Analysis.** The demo provides a dynamic visualization chart ("time-lapse") which allows the user to focus on a specific time interval and/or a subset of co-authors.
- **Ranking.** This tool also allows to compare and rank the overall independence of an author, along his/her whole career, with the whole research community (Figure 2). 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

1. M. Ausloos. A scientometrics law about co-authors and their ranking: the co-author core. *Scientometrics*, 95(3):895–909, 2013.
2. 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.
3. 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. 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.