It is not What but Who you Know: a Time-Sensitive Collaboration Impact Measure of Researchers in Surrounding Communities

Luigi Di Caro
University of Torino
Torino, Italy
dicaro@di.unito.it

Mario Cataldi
Université Paris 8
Paris, France
m.cataldi@iut.univ-paris8.fr

Myriam Lamolle
Université Paris 8
Paris, France
m.lamolle@iut.univ-paris8.fr

Claudio Schifanella
RAI research centre
Torino, Italy
claudio.schifanella@rai.it

ABSTRACT

In the last decades, many measures and metrics have been proposed with the goal of automatically providing quantitative rather than qualitative indications over researchers' academic productions. However, when evaluating a researcher, most of the commonly-applied measures do not consider one of the key aspect of every research work: the collaborations among researchers and, more specifically, the impact that each co-author has on the scientific production of another. In fact, in an evaluation process, some co-authored works can unconditionally favor researchers working in competitive research environments surrounded by experts able to lead high-quality research projects, where state-of-the-art measures usually fail in trying to distinguish co-authors from their pure publication history. In the light of this, instead of focusing on a pure quantitative/qualitative evaluation of curricula, we propose a novel temporal model for formalizing and estimating the *dependence* of a researcher on individual collaborations, over time, in surrounding communities. We then implemented and evaluated our model with a set of experiments on real case scenarios and through an extensive user study.

1. INTRODUCTION

Considering the research community, one of the most commonly adopted method to evaluate the career of a researcher is to consider his/her authored papers and evaluate their "impact" on the surrounding research community. But how to evaluate this impact is still debated: most of the existing methods rely on counting the number of papers co-authored by the researcher and/or estimating, by applying different approaches, their citations number and quality.

Although these measures represent valuable tools for analyzing researchers outputs, they usually assume the co-authorship to be a proportional collaboration among the involved parts, missing out their relationships and their relative scientific impact on the resulting work. Moreover, considering that many of these measures are also used for recruitment purposes, it could be crucial to analyze the scientific relationships among authors in order to estimate the capacity of an author to work and produce research outcomes *without* the people that assisted his/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 (and more demanded on recruitment processes) 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.

With this goal, we propose a novel temporal model that aims at evaluating the scientific collaborations of an author, over time, and their impacts on his/her entire research production (intended as the set of papers co-authored by him/her). Moreover, based on the DBLP bibliographic database[1], we also developed a web environment (`http://d-index.di.unito.it/`) that implements the presented model and proposes a set of visualization tools to permit to analyze, study and compare the careers of all the indexed authors based on their entire bibliographic records. Finally, relying on this platform, we present case and user studies that test both the validity and the reliability of the proposed evaluation measure.

WWW 2015 Companion, May 18–22, 2015, Florence, Italy.
ACM 978-1-4503-3473-0/15/05.
http://dx.doi.org/10.1145/2740908.2742023.

[1]http://dblp.uni-trier.de/db

2. RELATED WORK

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 [16], for example, the author first pointed out the problem of *undeserved coauthorship*. In [7] 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 [8]. 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 [5] and later in [6], the author introduced the h-index, a well-known metric for evaluations of academic careers, impact of journals, and research communities. An author has index h if he/she published h papers having at least h citations. As it has been fully demonstrated in [15], it indirectly measures both quality and quantity. Since the h-index has been introduced, several extensions have been studied to avoid its drawbacks. The *g(m)* index [14], for instance, counts the papers equally fractionally according to the number of authors. In [4] the author gives a credit to a co-authorship based also on the number of received citations.

Other works presented interesting ideas and insights on such a complex and multi-faceted domain. The authors of [1], for instance, stated that an author's scientific relevance should not be based on the number of citations of her/his papers, but is about how much co-workers she/he has been able to connect to in order to produce (joint) scientific publications. In [18] the authors started from the same motivation and proposed the *independence indicator* made up of three different dimensions of independence: the ability of developing own co-author networks, novel thematic directions, and strong quality of the research focus.

Many works tried also to take into account implicit and/or explicit edges of the collaboration network for detection and/or evaluation purposes. In [10] and [12], the authors aim at detecting characteristics like academic department, position, and country of origin from socio-academic networks, while [17] focus on the evolution of research teams. In [20], the authors use four centrality measures within a restricted collaboration network, showing that they are significantly correlated with citation counts. [11] aims at discovering the diffusion of scientific credits in the community relying on a citation network.

In [13, 2], we presented a novel approach to estimate the dependence of an author on a collaboration. In this paper, we further extended the work in order to analyze dependences over time and take into account how each single collaboration evolves in different time range (while the original works only considered static conditions).

3. MATERIALS AND METHODS

In literature there is plenty of methods for evaluating the output of an author (either called scientist or researcher in the paper). Most of them consider their publication records as the basis for their scientific evaluation. In our paper, given an author a_i, we formalize his/her set of research outputs (also called papers, works or outcomes from now on) $O^t_{a_i}$, published until the time t, as[2]

$$O^t_{a_i} = \{o^t_{a_i,1}, o^t_{a_i,2}, \cdots o^t_{a_i,n}\}, \tag{1}$$

where $o^t_{a_i,k}$ is the k-th research output authored, or co-authored, by him/her at the time t (for example, if $O^t_{a_i}$, with t=2000, contains all papers authored by a_i from the beginning of his/her career until 2000). Considering this information, it is possible to quantify the "productivity" of a_i, $p^t_{a_i}$, at the time t, as

$$p^t_{a_i} = |O^t_{a_i}|, \tag{2}$$

where $|O^t_{a_i}|$ is the cardinality of $O^t_{a_i}$.

In the same way, we can define the common outcome $O^t_{a_i,a_j}$, at the time t, of two authors, a_i and a_j, as

$$O^t_{a_i,a_j} = O^t_{a_i} \cap O^t_{a_j} = \{o^t_{(a_i,a_j),1}, o^t_{(a_i,a_j),2}, \cdots, o^t_{(a_i,a_j),m}\}, \tag{3}$$

where $o^t_{(a_i,a_j),k}$ is the k-th research output co-authored by both a_i and a_j at the time t, and m is the total number of papers co-authored by both of them at t. It is then possible to quantify their productivity $p^t_{a_i,a_j}$ at the time t as

$$p^t_{a_i,a_j} = |O^t_{a_i,a_j}|. \tag{4}$$

Notice that this approach is extendable to any set of authors with any cardinality.

Moreover, given an author a_i, we formalize the scientific network in which he/she produced research outcomes as

$$Net^t_{a_i} = \{a^t_1, a^t_2, \cdots, a^t_h\} \tag{5}$$

where h is the total number of co-authors, at the time t, of a_i. In the same way, given two authors a_i and a_j, we formalize their common scientific co-authorship network $Net^t_{a_i,a_j}$, at the time t, as the set of authors who co-authored at least one paper with both a_i and a_j, in the same output (i.e. $a_k \in Net^t_{a_i,a_j} \Rightarrow \exists o^t_x \in O^t_{a_i,a_j}$ s.t. o^t_x is also co-authored by a_k).

In the next sections we will leverage these formalizations to introduce our temporal model.

3.1 The d-index: Analyzing Dependences in Collaborative Environments

Given the entire set of scientific authors, theoretically, it is now possible to model a temporal co-authorship network, N^t, as a directed graph that expresses the dependence of each author, at the time t, on the scientific collaboration with a co-author. Formally we define N^t as

$$N^t = \{V^t, E^t, d\}, \tag{6}$$

where

- $V^t = \{a^t_1, a^t_2, \cdots, a^t_n\}$ is the complete set of n scientific authors at the time t (i.e. researcher having published at least one outcome at the time t);

[2] In this paper, the considered time intervals represent publication years.

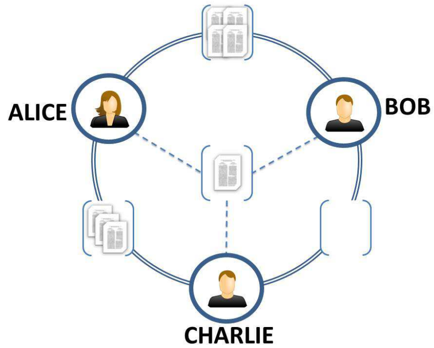

Figure 1: A simplified example of co-authorship relations: three scientists, *Alice*, *Bob* and *Charlie* published, all together, one paper. However, *Bob* and *Charlie* did not publish any scientific outcome without *Alice*.

- E^t is the set of undirected edges, where each $e_{i,j} \in E$ represents an existing collaboration at the time t between a_i and a_j (where $a_i, a_j \in V$) motivated by at least one research output co-authored by both of them at the time t;
- d is the weighting function ($d : E \rightarrow [0; +1]$) representing the dependence, at the time t of a_i on the scientific collaboration with a_j.

Following this formalization, in order to measure this dependence value, called d-index, we first aim to study each scientific collaboration of an author and estimate his/her autonomy from the surrounding scientific communities. Then, we will try to quantify the overall dependence of a considered researcher on the scientific collaboration with a specific co-author by analyzing how much each collaboration of the author were autonomous from the contribution of the considered co-author. In a sense, we aim at quantifying the impact of an author on the career of another by analyzing his/her average impact on all the his/her scientific collaborations.

In order to better understand the problem, let us consider a simplified situation as the one shown in Figure 1. Three authors, *Alice*, *Bob* and *Charlie* collaborated by publishing several scientific works. In particular, the collaboration between *Alice* and *Bob* (without *Charlie*) resulted in many research outputs as for the collaboration between *Alice* and *Charlie* (without *Bob*). On the other hand, the scientific relationships without *Alice* did not result in any published work. This situation can be summarized as follows: *Bob* and *Charlie* can be thought as young researchers who are supervised by *Alice*. In this case, *Alice* is leading this research group and most of the necessary expertise can be easily credited to her. This fact does not reduce the merits or the contribution of *Bob* and *Charlie* in the considered research outputs. We just state that this situation suggests that the scientific production of *Bob* and *Charlie* results highly dependent on the scientific collaboration with *Alice*, which is also confirmed by the fact that, for each of them, any collaboration without *Alice* resulted poorly productive.

Considering this example, the scientific dependence has been highlighted from the analysis of their *co-authorship network* that models the environment (and, therefore, the relationships existing among authors) in which they work. In a sense, through this model, we analyze the productivity and the autonomy of each collaboration with respect to all their co-authors and understand the impact of each author on the scientific production of each scientist in this collaborative environment.

Based on these assumptions, we first introduce a method to measure the "autonomy" of a collaboration, by taking into account the common scientific production, at the time t, of the involved authors. At this point, given two authors a_i, a_j and their common scientific network $Net^t_{a_i,a_j}$, the autonomy of their collaboration $w^t_{a_i,a_j}$ is calculated as

$$w^t_{a_i,a_j} = \begin{cases} 0 & \text{if } Net^t_{a_i,a_j} = \emptyset \\ \dfrac{1}{\sum_{a_k \in Net^t_{a_i,a_j}} \left(\sum_{x=1}^{c(a_k, O^t_{a_i,a_j})} \frac{1}{x} \right)} & \text{if } Net^t_{a_i,a_j} \neq \emptyset \end{cases} \tag{7}$$

where the function $c(a_k, O^t_{a_i,a_j})$ returns the number of times the author a_k co-authored a paper with both a_i and a_j at the time t. Intuitively, this formula permits to measure the independence, at the time t, of the collaboration between a_i and a_j from the collaboration with any other author of the common scientific environment, expressed by $Net^t_{a_i,a_j}$. In this way, we take into account number and frequency of each collaboration; from one side, we count how many external co-authors, along their collaboration history (until the time t), have been involved in the collaboration between a_i and a_j. From the other side, we also aim at evaluating the frequency of each contribution on their collaborations. In a sense, the autonomy of the collaboration will be lower when a high number of external co-authors are repetitively involved in the scientific outputs of the collaboration. Intuitively, the higher the autonomy, the more independent the work of a_i and a_j from the collaboration with any other co-author (and the other way around).

From this, given an author a_i, we aim at calculating his/her overall dependence on the collaboration with a_j by taking into account the capacity of a_i of working in his/her scientific environment without the scientific support of a_j. For this, given an author a_i and his/her scientific environment $Net^t_{a_i}$, at the time t, we define the dependence value, d-index, of the co-author a_i on the collaboration with a_j as $d^t_{a_i \rightarrow a_j}$

$$d^t_{a_i \rightarrow a_j} = \frac{p^t_{a_i,a_j}}{p^t_{a_i}} \times \frac{w^t_{a_i,a_j,Net^t_{a_i}} + w^t_{a_j,\neg a_i,Net^t_{a_i}}}{w^t_{a_i,a_j,Net^t_{a_i}} + w^t_{a_j,\neg a_i,Net^t_{a_i}} + w^t_{a_i,\neg a_j,Net^t_{a_i}}}, \tag{8}$$

where

- $p^t_{a_i}$ returns the productivity of a_i at the time t;
- $p^t_{a_i,a_j}$ is the productivity of the collaboration between a_i, a_j at the time t;
- $w^t_{a_i,a_j,Net^t_{a_i}}$ is the autonomy of the collaboration, at the time t, among a_i, a_j and $Net^t_{a_i}$ (i.e. the autonomy score of the collaboration between a_i and a_j, and at least one author a_k in $Net^t_{a_i}$);
- $w^t_{a_i,\neg a_j,Net^t_{a_i}}$ is the autonomy score of the collaboration between at least one author in $Net^t_{a_i}$ and a_i with-

out the contribution of a_j (i.e., excluding the research outputs in which a_j is also involved);

- $w^t_{a_j,\neg a_i,Net^t_{a_i}}$ is the autonomy score of of the collaboration between a least one author in $Net^t_{a_i}$ and a_j without the contribution of a_i (i.e., excluding the research outputs in which a_i is also involved).

The d-index value $d^t_{a_j \to a_i}$ ranges from *0* to *1*; in particular, $d^t_{a_i \to a_j} \approx 0$ indicates that the dependence of a_i on a_j, at the time t, is negligible, while a $d^t_{a_i \to a_j} \approx 1$ highlights the contrary. In fact the second term of the formula increases when the autonomy score of a_i and $Net^t_{a_i}$, without the contribution of a_j, is negligible ($w^t_{a_i,\neg a_j,Net^t_{a_i}} \approx 0$) and the other collaborations are significantly autonomous ($w^t_{a_i,a_j,Net^t_{a_i}} > 0$ and $w^t_{a_i,\neg a_j,Net^t_{a_i}} > 0$). On the other hand, the higher the $w^t_{a_i,\neg a_j,Net^t_{a_i}}$, the lower the relative dependence.

Please also notice that $d^t_{a_i \to a_j} \neq d^t_{a_j \to a_i}$; in fact their mutual dependences can significantly differ, since they are also based on their personal collaborations (which are obviously not the same, even when they share the same co-authors).

3.2 Dependence Trajectory: Leveraging the d-index Values to Estimate the Evolution of the Dependence over Time

In Section 3.1, we introduced a novel way to estimate the dependence, at a specific time, of a given author on the scientific collaboration with a co-author, based on their scientific network and the productivity of each collaboration within this network. These values can now be leveraged to graphically map the scientific dependences of an author, along his/her career, on the collaboration with each co-author, as a set of curves that plots the relative d-index values. For this, we define the dependence curve of an author a_i with respect to a co-author a_j as

$$\overrightarrow{d_{a_i \to a_j}} = \{d^t_{a_i \to a_j}, d^{t+1}_{a_i \to a_j}, \cdots, d^{t+n}_{a_i \to a_j}\}, \tag{9}$$

where t is the year of the first publication of a_i, and n expresses the arithmetical difference between the last and the first year of publication of a_i. Thus, given an author a_i, and the complete set of his/her coauthors expressed by Net_{a_i}, it is now possible to graphically represent, in the same chart, his/her dependence on each co-author $a_k \in Net_{a_i}$, along the career of a_i, to obtain a first sight on this mined knowledge. An example is shown in Figure 2 (a).

Each of these curves can graphically highlight the evolution of the collaboration with a specific co-author along the time and understand how much the considered author became independent (or dependent) from him/her with the years. Considering the example in 2 (a), nine dependence curves are provided. Eight of them are visibly decreasing (highlighting the fact that the author becomes increasingly independent from the collaboration with the eight related co-authors, along the considered intervals) while the last one, in red, significantly increases after 2005 (thus, even if the number of co-authors increases, the author becomes dependent on the collaboration with the related co-author).

In order to better evaluate this situation, considering that many authors can have hundreds of co-authors on a career of multiple decades, in this section we also aim at obtaining

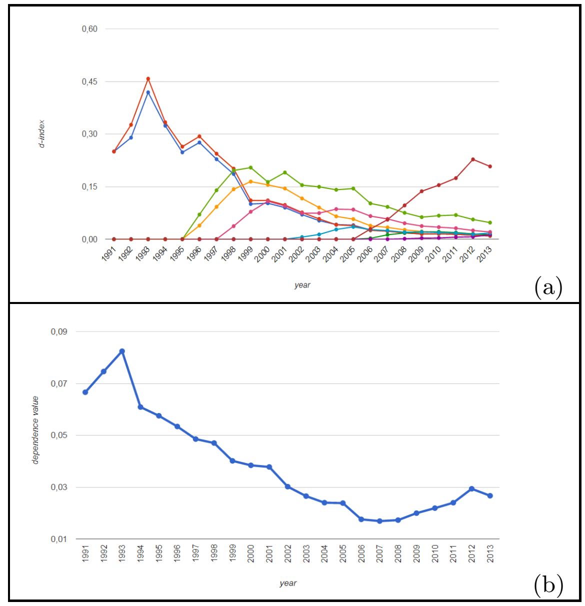

Figure 2: (a) The dependence curves of an author a_i and his/her dependence trajectory (b).

a one-curve evaluation system to summarize, at best, the overall independence of the author.

Thus, given the complete set of dependence-curves, we calculate the author's *dependence trajectory*, by calculating the standard deviation, along the time, of each d-index 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 author is independent from the collaboration with the considered co-author). In a sense, we aim at evaluating the overall independence of an author from the surrounding community. More formally, given an author a_i, we define his/her dependence trajectory $\overrightarrow{d_{a_i}}$ as

$$\overrightarrow{d_{a_i}} = \{sd^t_{a_i}, sd^{t+1}_{a_i}, \cdots, sd^{t+n}_{a_i}\}, \tag{10}$$

where $sd^t_{a_i}$ is calculated as

$$sd^t_{a_i} = \sqrt{\frac{\sum_{a_k \in Net_{a_i}} (d^t_{a_i \to a_k})^2}{|Net_{a_i}|}}. \tag{11}$$

The meaning of this formula is evident: calculate the average standard deviation of the previously calculated d-index values from the optimal value of 0. The higher the $sd^t_{a_i}$, the more dependent the work of a_i on the collaboration with any of his/her co-authors at the time t.

In Figure 2 (b) we show an example of dependence trajectory (calculated based on the dependence curves shown in Figure 2 (a)). In this example, we can easily see an overall increment in the dependence trajectory; this is mainly due to the significant increase in the dependence values of the considered author related to a specific co-author (visualized through the red line in Figure 2 (a)). The reason of this behavior is evident: the system tries to detect anomalies in the collaboration patterns with respect to some expected values. Authors in fact are expected, along the career, to increment their collaboration network and, therefore, become independent from the collaboration with each single co-author. Equation 7 in fact leverages the number of col-

Position	#papers	#co-authors	#users
Ph.D Stud/Post Doc.	19.28	26.21	24
Res./Assist. Prof.	31.57	29.04	29
Professors	64.43	45.93	28

Table 1: The number of users that replied to our questionnaire, grouped based on their academic position.

laborations to estimate the autonomy of a scientific relationship. However, in the considered example, even in presence of a general (expected) increase in the number of scientific collaborations along the career, the increment in the dependence on a single co-author is so significant to lead the system to a visible boost in the dependence trajectory (which, however, is expected to constantly decrease).

In the next sections we will show how to use this evaluation system for analysis and comparison purposes.

4. RESULTS

In this paper we have presented a novel temporal model that permits to focus the evaluation of an author on the analysis of his/her scientific collaborations. The model has been implemented and all the data are freely accessible at *http://d-index.di.unito.it*.

For our experiments, we considered a data-set extracted from the DBLP bibliographic database[3] containing information about 1,342,723 authors and 2,446,236 scientific papers.

In this section, we illustrate the results of a user study that we conducted to evaluate the impact of academic collaborations, over time, on the career of the researchers. A detailed view of the users set is shown in Table 1.

In detail, we asked users to answer a questionnaire in which they had to order three randomly-picked co-authors according to who they felt being more important in her/his career. We repeated this question 6 times, i.e. 2 times for 3 different time frames, randomly picked within the beginning, middle and final time intervals of the career. We define the beginning of a career as the period included within the first third of his/her publication history, the middle as the second third and the end as the third third of the career of an author. We also added two check-boxes called "*difficult choice*" to let the users express a possible doubt between the first-ranked co-author and the second-ranked co-author, and/or between the second-ranked and the third-ranked respectively. This option ensured the possibility for the user to express doubtful responses.

4.1 Validation and Experiments

In order to asses the validity of the approach, we compared the users rankings with the ones provided by our system through the d-index values. This comparison has been first performed by means of the well-known Pearson product-moment correlation coefficient r, which is the covariance of the two variables divided by the product of their standard deviations. The correlation coefficient ranges from -1 to 1 (where 0 means independence), and in our test we achieved a total r score of 0.76. Please notice that values greater 0.7 indicate a strong correlation [3].

Since this coefficient is only able to capture linear correlations, and it is known to be not robust especially in case of

[3]http://www.informatik.uni-trier.de/ley/db/

Time Range	KT	WKT	KT (DC)	WKT (DC)
Early career	0.561	0.648	0.645	**0.704**
Mid-career	0.556	0.685	0.634	**0.741**
End career	0.536	0.668	0.552	**0.662**
Overall	0.551	0.667	0.610	**0.702**

Table 2: The results of the user study in terms of both original and weighted Kendall's Tau coefficient. The term *(DC)* indicates that the difficult choices marked by the users have not been considered.

outliers [19], we further evaluated our d-index values making use of the Kendall's Tau coefficient [9]. Using this measure it is possible to capture the rank correlation between different orderings. In detail, this measure works as follows: given a set of non-ordered items $I = \{i_1, i_2, \cdots i_n\}$ (in our case, representing co-authors), two ordered sets R_a and R_b of all the elements in I (i.e., the orderings provided by the user and the system), and the set T of all the ordered pairs of elements in I, $< i_p, i_q >$ (where $p > q$) , the Kendall's Tau coefficient calculates the distance between the two orderings by relying on the number of pairs in T that are concordant and discordant in the considered rankings. More formally, the Kendall's Tau coefficient is calculated as:

$$kendall(R_a, R_b) = \frac{conc - disc}{|T|} \tag{12}$$

where $|T|$ is the number of the pairs, $conc$ is the number of pairs in T that are equally ordered within R_a and R_b, and $disc$ is the number of pairs in T that are differently ordered within R_a and R_b. Please notice that Kendall's Tau coefficient ranges from -1 (one ordering is the contrary of the other) to +1 (the considered orderings are exactly equals).

The results are shown in Table 2, aggregated and averaged by different time frames. These experiments clearly highlight both the validity and the reliability of our approach: the resulting Kendall's Tau coefficients result very coherent for all the aggregations and the considered time intervals. The Kendall's Tau coefficients are indeed similar for all the considered academic intervals (0.561 for Early career, 0.556 for mid-career and 0.536 for the end of the career), therefore proving the capacity of the proposed approach in positively capturing collaboration dependences over time.

Moreover, in order to better evaluate the approach, we have slightly modified the reported Kendall's Tau definition in order to take into account the weighted distances provided by the ordered d-index values. We believe that a comparison among users and system orderings should take into account also the relative distances, in terms of d-index values, among the considered co-authors. In other words, in the Kendall's Tau computation, the presence of a discordant pair with a high d-index distance between the two items should have an higher negative impact with respect to a discordant ordering with similar d-index values, and the other way around. More formally, we then computed these values as

$$conc = \sum_{<i_p, i_q> \in T} \left(1 - \frac{dist(i_p, i_q)}{max(I) - min(I)}\right) \tag{13}$$

and

$$disc = \sum_{<i_p, i_q> \in T} \left(\frac{dist(i_p, i_q)}{max(I) - min(I)} \right) \quad (14)$$

where $dist(i_p, i_q)$ is the distance between the d-index value of i_p and the one of i_q (calculated as the absolute value of the difference), while $max(I)$ and $min(I)$ are respectively the highest and the lowest d-index values related to some item in I. This way, we are able to weight the differences between system and users orderings accordingly to the d-index values (i.e., the higher the difference between two d-index values the more significant the correct matching between the ranks, and vice-versa). Even with this weighted normalization, the Kendall's Tau coefficient still ranges between -1 and +1. Again, the results shown in the "WKT" column in Table 2 demonstrate that all the Kendall's Tau coefficients increase when the d-index distances are taken into account, highlighting that the system is able to capture differences among scientific dependencies where they matter while it can fail mostly when they are minimal.

Finally, we thought that making decisions between collaborators in terms of their scientific dependence could be difficult in some cases. For this reason, as already reported, the users had the possibility to mark those choices that they felt difficult to take (check-box "*difficult choice*" in the User Study). Table 2 shows the results of the evaluation where these doubtful answers are not taken into account (*DC* columns). At this point, considering users' difficult choices and weighting the rankings with respect to the obtained d-index values, we reached an overall positive Weighted Kendall's Tau score of 0.702, which highlights the capacity of the proposed model to capture the dependences where they have an evident impact on a research curriculum. To sum up, the d-index resulted to achieve high positive correlation values with the users feelings. Moreover, we also demonstrated that the approach is also able to capture the concept of scientific impact of collaboration over time, i.e., in different time frames. An interesting insight is about the light decrease of the Kendall's Tau scores in the last years of the authors' career, probably due to the incremental complexity of the overall collaboration networks.

5. CONCLUSIONS

The problem of evaluating the quality of researchers' outputs has been broadly studied whereas few works attempted to discover collaboration dependences among researchers in case of co-authored papers. In this sense, we proposed a novel temporal model that aims at uncovering dependences among the authors over time according to their research environment and their publication history. We then evaluated the presented model through several examples and user studies that validated the model under different points of view. We also introduced the freely available web platform `http://d-index.di.unito.it/` that implements the presented ideas.

6. 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] C. P. Dancey and J. Reidy. *Statistics without maths for psychology*. Pearson Education, 2007.

[4] L. Egghe. Mathematical theory of the h-and g-index in case of fractional counting of authorship. *Journal of the American Society for Information Science and Technology*, 59(10):1608–1616, 2008.

[5] J. 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, 2005.

[6] J. Hirsch. Does the h index have predictive power? *Proceedings of the National Academy of Sciences*, 104(49):19193, 2007.

[7] R. Hunt. Trying an authorship index. *Nature*, 352(6332):187–187, 1991.

[8] J. Imperial and A. Rodríguez-Navarro. Usefulness of Hirsch's h-index to evaluate scientific research in Spain. *Scientometrics*, 71(2):271–282, 2007.

[9] M. Kendall. A new measure of rank correlation. *Biometrika*, 30(1/2):81–93, 1938.

[10] A. Pepe and M. Rodriguez. An in-depth longitudinal analysis of mixing patterns in a small scientific collaboration network. *Scientometrics*, 85(3), 2010.

[11] F. Radicchi, B. Markines, and A. Vespignani. Diffusion of scientific credits and the ranking of scientists. *Physical Review E*, 80(5):056103, 2009.

[12] M. Rodriguez and A. Pepe. On the relationship between the structural and socioacademic communities of a coauthorship network. *Journal of Informetrics*, 2(3):195–201, 2008.

[13] C. Schifanella, L. D. Caro, M. Cataldi, and M. Aufaure. D-INDEX: a web environment for analyzing dependences among scientific collaborators. In *The SIGKDD'12*, pages 1520–1523, 2012.

[14] M. Schreiber. To share the fame in a fair way, hm modifies h for multi-authored manuscripts. *New Journal of Physics*, 10(4):040201, 2008.

[15] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos. Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics*, 72(2):253–280, 2007.

[16] 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.

[17] C. Taramasco, J. Cointet, and C. Roth. Academic team formation as evolving hypergraphs. *Scientometrics*, 85(3):721–740, 2010.

[18] P. Van den Besselaar, U. Sandström, and I. Van der Weijden. The independence indicator: Towards bibliometric quality indicators at the individual level. 2012.

[19] R. R. Wilcox. *Introduction to robust estimation and hypothesis testing*. Academic Press, 2012.

[20] E. Yan and Y. Ding. Applying centrality measures to impact analysis: A coauthorship network analysis. *Journal of the American Society for Information Science and Technology*, 60(10):2107–2118, 2009.