The Science of Science and a Multilayer Network Approach to Scientists' Ranking

Georgios Sideris
University of Thessaly
Volos, Greece
sideris@uth.gr

Dimitrios Katsaros
University of Thessaly
Volos, Greece
dkatsar@inf.uth.gr

Antonis Sidiropoulos
Alexander Technological Educational Institute of Thessaloniki
Thessaloniki, Greece
asidirop@it.teithe.gr

Yannis Manolopoulos
Open University of Cyprus
Nicosia, Cyprus
yannis.manolopoulos@ouc.ac.cy

ABSTRACT

The deluge of data on scholarly output created unique opportunities for identifying the drivers of modern science, for studying career paths of scientists, and for measuring the research performance. These massive data and processing methodologies have given rise to an exciting new field, namely Science of Science (SoS) as the successor of what is called scientometrics or informetrics for many decades. Science of Science is the offspring of the fertile cooperation of many disciplines, such as network science, statistics, machine learning, mathematical analysis, sociology of science and so on. In this article, we provide a comprehensive coverage of recent advances in SoS related to network analysis, prediction and ranking, and investigate the issue of scientist ranking from a multilayer network perspective. Towards this goal, we contrast by experiments the well-known h-index and the recently proposed indicator C^3-index to a generalization of PageRank for multilayer networks, namely BiPlex PageRank, which is based on solid tensor analysis. Both the obtained results and the brief survey of SoS will deepen our faith to SoS and stimulate further efforts in this transdisciplinary field.

CCS CONCEPTS

• **General and reference → Metrics**; **Evaluation**; • **Human-centered computing → Collaborative and social computing design and evaluation methods**; **Social network analysis**; *Social engineering (social sciences)*;

KEYWORDS

Scientometrics, Bibliometrics, Ranking, Evaluation, Informetrics, Science of Science, SoS

IDEAS 2018, June 18–20, 2018, Villa San Giovanni, Italy

ACM ISBN 978-1-4503-6527-7/18/06...$15.00
https://doi.org/10.1145/3216122.3229862

ACM Reference Format:
Georgios Sideris, Dimitrios Katsaros, Antonis Sidiropoulos, and Yannis Manolopoulos. 2018. The Science of Science and a Multilayer Network Approach to Scientists' Ranking. In *IDEAS 2018: 22nd International Database Engineering & Applications Symposium, June 18–20, 2018, Villa San Giovanni, Italy.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3216122.3229862

1 INTRODUCTION TO THE SCIENCE OF SCIENCE

Modern science evolves in a much rapid and different way that it used to. Moreover, in our extremely digitized world we now have an abundance of data that document this evolution, and many ways to disseminate knowledge and collaborate by breaking the barrier of geographic isolation. Thus, recently the so-called Science of Science[10, 38] (SoS) is growing at an unprecedented pace. SoS seeks to understand, quantify and even predict research activities and the corresponding outcomes. The first factor that enabled the emergence of SoS is the data availability provided – freely many times – by sources such as Scopus, Google Scholar, Web of Science, PubMed, Microsoft Academic and so on. The second enabling factor was the extensive collaboration of scientists, and the exchange and osmosis of ideas coming from many (seemingly) diverse disciplines, such as network science, scientometrics, sociology and so on. SoS's scope is certainly broader than Scientometrics; the latter studies the impact of publications, researchers, journals and tries to model the scientific collaboration, to establish concrete comprehension of innovation and predict future evolution of science. SoS uses/develops models "to probe the mechanisms driving science, from knowledge production to scientific impact" [38]. What actually drives the development of SoS is the industrialization of science in almost every aspects of it, from scientific impact recording, to research article reading/citing recommendation, to hiring/promotion decisions, and to research fund allocation.

SoS was mainly benefitted from the advances in what is termed *complexity science* and now it is mostly known as *network science* [2, 19, 22]. Complexity science studies the complex systems that consist of many entities and interact in non-trivial ways, such that looking at the individual element behaviour cannot provide explanations for the behaviour of the whole system. Since a complex system can be modeled as a graph (nodes are the constituent entities, whereas links represent the interactions among pairs of elements) it is commonly perceived as being the same as network science, which studies modern networks and their properties. Even though this is an erroneous perception, for the sake of simplicity and brevity in this short introduction we will accept it. Thus, network science offered to SoS network (growth) models such as the small world [35] and scale-free [3], introduction of centrality measures [22], and the study of diffusion processes such as influential spreaders [5, 16], influence maximization [15] and so on. Moreover, network science introduced novel network types with richer modelling capabilities, where the interacting entities are assumed to belong to more than one network, called *layer*. These networks are called multiplex [8], multisliced [21], multilevel [34], interdepedent [7], or in general multilayer [6, 17]. The influence of network science on SoS methodologies is explained by the fact that contemporary science is a networked dynamical system among social structures (e.g., scientists), knowledge representation (e.g., research articles) and the nature world that are interacting in a very complex way.

Trying to provide a synopsis of what is SoS about then we would list the following topics:

- (Static) Data: scientific publication data, funding data, patent data, collaboration network data, citation network data.
- (Dynamics of) publications/citations/collaborations: preferential attachment growth, aging effect, sleeping beauties, team formation/assembly.
- Scientific significance: citation measures, spectral centralities measures, scientist/journal/university/country ranking, credit allocation.
- Prediction in science: link prediction, future impact prediction.
- Success in science: interdisciplinary research, funding, collaborators.
- Innovation and knowledge diffusion: patents, coauthorships, researchers mobility.

In the rest of this paper we will mainly focus on the third item, and provide a brief coverage of the fourth one as well. The structure of the paper is as follows: Section 2 will briefly cover the issues and advances in quantifying scientific significance 2; section 3 will survey the topic predictive modeling in science; section 4 will present details of three scientific ranking indicators for multilayer scientific networks, whereas section 6 will present an evaluation of them. Finally, section 7 will conclude the article.

2 QUANTIFYING SCIENTIFIC SIGNIFICANCE

Quantification of scientific performance and significance concerns individual articles, scientists, journals, institutions, and whole countries. The great majority of methods to quantify the impact of these entities is to reduce into quantifying the impact of some articles that somehow are related to the entity (e.g., authored by a specific scientist, appearing in a specific journal, and so on). Thus, we will focus mainly on article and then on scientist impact quantification. A straightforward way to measure impact is by citation counting. However, citing behavior varies across disciplines; thus, other methods based on spectral centralities have appeared in the literature. The most widely known such method is PageRank [18] and its variants[1], e.g., CiteRank, DivRank, PrestigeRank, NonlinearRank, SRank etc.

Traditional measures for scientists' performance quantification include the number of published articles, number of citations received, average number of citations per published article, and so on. A path-breaking idea was the introduction of the h-index [13] which is actually a proxy for both productivity and impact. The introduction of the h-index spawned a large number of variants, among them our own contemporary index [30]. A different line of research started from the observation that author centrality in collaboration networks is strongly correlated to author impact; thus these works proposed the estimation of author impact by applying variations of PageRank such as the AuthorRank [20], or personalized version of the PageRank in citation networks. Despite the very rich litarature on impact indicators for scientists using citation data from plain citation networks, only a handful of works addressed the same issue in multilayer networks. An (unvalidated) work described the AuthorPaper rank (APrank) index [42], which – similar to the idea of Hyperlinked Induced Topic Search (HITS) algorithm [18] – interweaved paper and author quality into a recursive definition: a paper is of high quality if it is cited by prestigious scientists and that high-quality papers raise the prestige of their authors. Another notable effort is the P-Rank [37]. It is interesting to note however that none of these methods employed native spectral centralities for multilayer networks as impact indicators.

3 PREDICTIONS IN SCIENCE

The most common prediction problem is that of forecasting future performance of an entity, either this is a research article or a scientist. A second interesting prediction problem is that of predicting future links in collaboration networks or identifying missing links in citation networks. In the next two subsections we elaborate on them.

3.1 Impact prediction

Paper impact prediction has (almost exclusively) focused on citation prediction. Since a citation network is a growing network, theories developed in the context of complex network growth were the most promising for addressing this problem. Thus, the celebrated preferential attachment model along with linear extrapolation was applied [39] for citation prediction. The accuracy of that model was good for the short term, because fast aging of papers and substantial diversity in their quality turned it inadequate for long term predictions. In a path-breaking paper [33], a preferential attachment model along with fitness and aging parameters was introduced that exhibited extraordinary accuracy.

Scientist impact prediction is a much more complicated issue because of the vast number of performance indicators proposed in the literature, even though most of them are based on the standard impact measure, i.e., citations. A very interesting line of work is the task of predicting future h-index and was put forward in [1], which deployed a linear regression model and established a set of equations (Equations 1–3) for predicting h-index value for one, five and ten years ahead, based on the following parameters: n is the

[1]For the interest of space, we refrained from providing the citations to the articles introducing these methods; instead the reader can find it in [38].

number of articles written by the author, h is the current h-index of the author, y is years since the author published the first article; j is the number of distinct journals in which the author has published, and q is the number of the author's top journal articles.

$$h_{+1} = 0.76 + 0.37\sqrt{n} + 0.97h - 0.07y + 0.02j + 0.03q \quad (1)$$

$$h_{+5} = 4 + 1.58\sqrt{n} + 0.86h - 0.35y + 0.06j + 0.2q \quad (2)$$

$$h_{+10} = 8.73 + 1.33\sqrt{n} + 0.48h - 0.41y + 0.52j + 0.82q \quad (3)$$

However, the validity of these equations was questioned by several works, e.g., [11, 26]. Later on, a more successful prediction model was proposed in [31] based on the model developed in [33]. However, all these works and the subsequent ones concluded that it is pretty difficult to make predictions for young scholars, which is somehow disappointing since this is exactly the main goal of the scientist's future impact prediction. More recent works focused on predicting rising stars [23], on incorporating the identify of early citers into the estimation of future impact [32], or providing more elaborate machine learning models for performance prediction [36].

3.2 Link prediction

The essence of any link prediction problem is the identification of similar not yet connected nodes in the respective complex network. Therefore, the problem reduces to that of defining and then examining similarity among nodes based on some topological and other features of the nodes. The idea is that similar nodes are more probable to be connected in the future, which is also supported by empirical evidence [19].

Some commonly used similarity measures in the context of social networks include the following:

- Number of common neighbors among the two examined nodes.
- Jaccard coefficient of the sets of neighbors of the two examined nodes.
- Resource allocation [41] which penalizes those neighbors of the nodes having large degree.
- Katz index [14] which calculates nodes similarity via paths counting.

Scientific collaboration networks are actually social networks, and thus all the aforementioned methods apply. However, recent work on graph embedding show more promising results compared to that traditional methods and metrics. The main procedure in a graph embedding is the assignment of a vector to every node and then embed this vector into some space, e.g., Euclidean, or hyperbolic [24]. Then, node similarity is computed based on the distance among vectors in that space. A representative work along this direction is the DeepWalk [27].

On the other hand, citation prediction is a different problem than prediction in social networks. There are two reasons for that. Firstly, citation prediction involves nodes that are going to be introduced into the network, i.e., new articles and thus social network link prediction methods do not apply because they work only for existing nodes; therefore, the problem becomes just the (uninteresting) case of identifying missing links in citations networks. The second reason is that citation networks are directed networks, contrary to social networks which are (usually) undirected. Therefore, some missing links prediction techniques were based on directed motifs [40], or others were based on applying machine learning techniques i.e., regression [29].

4 RANKING BASED ON MULTILAYER NETWORKS

In this section we provide a formal definition of multilayer networks and analyze briefly the inappropriateness of related ranking indices, which have been proposed for multilayer/multiplex networks.

4.1 Multiplex and multilayer networks

A *Single* or *Monoplex* network is represented as a graph $G_i(V_i,E_i)$, where V_i is the set of nodes and E_i is the set of edges that connect these nodes. Edges can be directed or undirected, weighted or unweighted. A *multilayer network* can be described as a combination of graphs, G_1, G_2, ..., $G_{|L|}$, and a set of interconnections between nodes in separate graphs. Edges connecting nodes of a single graph are featured as *intra-edges*, whereas edges connecting nodes of different graphs are notated as *inter-edges*. Formally, we describe a multilayer network as $\mathcal{P}(\mathcal{G},\mathcal{E})$, where $\mathcal{G} = \{G_i; i = 1, 2, .., |L|\}$ is a set of graphs, i.e., the layers of $\mathcal{P}$, and $\mathcal{E} = \{E_{ij} \subseteq V_i \times V_j;\ i, j \in \{1, 2, ..., |L|\},\ i \neq j\}$ is the set of inter-edges between nodes of different layers, i.e., different graphs. Figure 1 depicts a three layer multilayer complex network.

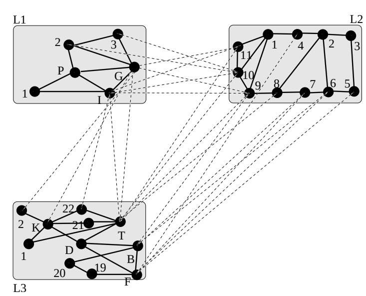

Figure 1: A multilayer network consisting of three layers L1, L2 and L3. Nodes with the same ID in different layers depict clones of the same node.

Multiplex networks are a special case of multilayer networks, where nodes are clones (counterparts) of themselves in each layer, i.e., $V_1 = V_2 = ... = V_{|L|} = V$. For multiplex networks the only inter-connections allowed are between a node and its counterparts in the remaining layers. Formally, $E_{ij} = \{(v, v);\ v \in V\}$ for all $i, j \in \{1, 2, ..., N\}$ with $i \neq j$.

4.2 Popular ranking indices for multiplex networks

In the context of multiplex networks there has been some work on generalizing PageRank. However, the generalizations provided are strictly for multiplex and not generic multilayer networks, and a straightforward adaptation of them to multilayer networks is not possible. We briefly present the most popular generalizations of PageRank for multilayer networks.

Additive PageRank for multiplex networks. In [12] the original PageRank algorithm is extended for multiplex networks requiring though a "predefined" ordering of the layers. For instance, in the so-called *additive Multiplex PageRank*, the effect of layer i on layer j is exerted by 'adding' some value to the centrality the nodes have in layer j in proportion to the centrality they have in layer i. Since there is no obvious ordering of the layers, the method is problematic.

Versatility PageRank. A fundamentally different flavor in extending PageRank for multiplex networks has been described in [9], which, using a tensorial notation, they provide a generalization of the original PageRank for multiplex networks, called the Versatility PageRank. Their method is very elegant and mathematically sound, but not directly applicable in generic multilayer networks.

Multilayer h-index. There has even been proposed an extension of the h-index to the context of multilayer networks [4], but it is not clear yet how to apply it to scientometric networks. Thus, in the next subsections we present two ranking methods that take into account layering information.

5 RANKING IN MULTILAYER NETS

In this section we present two elegant solutions for ranking scientists exploiting the multilayer structure of citation networks and also introduce a hybrid of them to reap their best features.

5.1 The C^3-index for scientist's ranking

The C^3-index was recently proposed in [28]. It represents a *non 'native'* index for multilayer networks in an attempt to embed information from both the citation network and the coauthorship network, and combine them into a single measure. However, its main motive was the fact that other indices, such as the h-index, produce great results in highly-cited scientists, but falls short in ranking medium and low cited ones. C^3 promises more accurate results for these types and identifying researchers with a promising future ahead of them. C^3-index combines three different measures and produces a ranking containing more information for each author. The three metrics are the following:

- ACI – Author citation Index
- PCI – Paper citation Index
- AAI – Author coAuthorship Index

We present next the equation that define the ranking.

$$C_j^3(t) = (1-\theta) + \theta \times (ACI_j(t) + AAI_j(t) + PCI_j(t)) \quad (4)$$

$$ACI_j(t) = (1-\theta) + \theta \times \sum_{A_k \in C(A_j)} \frac{ACI_j(t-1)}{outdeg(A_k)} \quad (5)$$

$$AAI_j(t) = \sum_{A_k \in C(A_j)} \frac{AAI_k(t-1)}{deg(A_k)} \quad (6)$$

$$PCI_j(t) = (C_j^3(t-1))^{\alpha} \times \sum_{P_k \in C(P_i)} \frac{PQI_k(t-1)}{\sum_{A_l \in A(P_k)} (C_l^3(t-1))^{\alpha}} \quad (7)$$

$$PQI_i(t) = (1-\theta) + \theta \times \sum_{P_k \in C(P_i)} \frac{PQI_k(t-1)}{outdeg(P_k)} \quad (8)$$

where $C(A_j)$ denotes the set of authors who cited at least one paper of author A_j, $CA(A_j)$ denotes the set of authors who coauthored with author A_j at least one paper, $outdeg(A_k)$ denotes the sum of the degrees of the outgoing edges from node A_k in the author-author citation layer of the network, $deg(A_k)$ denotes the sum of the degrees of the edges incident on node A_k in the author coauthorship layer, whereas θ is the well-known damping factor for PageRank. Also, t and $t-1$ represent times: t represents the current iteration's time, and $t-1$ the previous one's.

Thus, C^3-index is a PageRank-based multi-faceted metric for scientist's performance measurement, which combines the effect of citations and collaborations of an author in a systematic way using a weighted multi-layered network to rank authors.

5.2 The Biplex PageRank

Similar to Google's PageRank that it is used to rank Web pages, Biplex PageRank was proposed in [25] and extends the notion of PageRank to be applied to multilayered networks aiming to identify the most significant nodes from a spectral centrality perspective.

We will briefly provide the mathematics (Equations 9–13) that define the Biplex PageRank vector for a biplex network G with n nodes. P_a and P_{a_2} are the adjacency matrices of the two layers, respectively, α is the damping factor, and v is the personalization vector [18].

$$P = \frac{1}{2} \times (Pu + Pu_2 + Pd + Pd_2) \quad (9)$$

The calculation of Pu, Pu_2, Pd, and Pd_2 is performed iteratively according to the following equations:

$$2Pu^T = Pu^T * \alpha * Pa + Pu_2^T + 2\alpha Pd^T \quad (10)$$

$$2Pu_2^T = Pu^T + Pu_2^T * \alpha P_{a_2} + 2\alpha Pd_2^T \quad (11)$$

$$2Pd^T = (1-\alpha) * (Pu_2^T + Pd^T * e * v^T + Pd_2 * e * v^T) \quad (12)$$

$$2Pd_2^T = (1-\alpha) * (Pu_2^T + Pd^T * e * v_2^T + Pd_2 * e * v_2^T) \quad (13)$$

where Pu is the PageRank for the "real network", Pd is the PageRank for the "teleportation network", d is the vector of length n with "1s" in the place that corresponds to dangling nodes and "0s" otherwise, whereas u is the probability distribution of the dangling nodes.

The initial values of Pu, Pu_2, Pd, Pd_2 are for every element x of any of these vectors: $x = \frac{1}{2n}$, the damping factor is set to the usual value of 0.85, whereas the personalization vector is set to the uniform vector.

5.3 The C^4-index for scientist ranking

We introduce a new index, namely C^4-index aiming at identifying scientists with the most outstanding articles, while at the same time paying attention to their consistent focus on producing excellent results and also showing the biggest potential. To this end, we introduce a fourth layer to C^3-index's layers, the author significance layer but being produced by a multilayer centrality measure, in our case Biplex PageRank.

- ACI – Author citation Index
- PCI – Paper citation Index
- AAI – Author coAuthorship Index

- NBP – Author Centrality Index (Normalised Biplex PageRank)

The steps the describe C^4-index are the following:

- Calculation of the ACI, PCI , AAI for every author in the dataset (as done by C^3-index).
- Computation of Biplex PageRank
- Normalization of Biplex PageRank results
- Computation of $C_j^4(t)$ as in Equation 14

$$C_j^4(t) = (1-\theta)+\theta\times(ACI_j(t)+AAI_j(t)+PCI_j(t)+NBP_j(t)) \quad (14)$$

6 PERFORMANCE EVALUATION

6.1 Dataset and competitors

The main input for the tests applied on all algorithms is a dataset extracted from MAS (Microsoft Academic Search). The selected field was Computer Science and the top 500 authors were extracted according to MAS's ranking (and all others who had interacted with these top 500, e.g., being cited by them). The actual number of authors was 50601, who collectively published 13566 articles. Finally, there were 252142 citations in total. Three files were used of the dataset. The first one contained the coauthorship details of every article published by every scientist in the dataset, the second one contained the citations between any of the aforementioned articles, and the third one the names of the first 500 authors, for display purposes. Striving for anonymity, we refrained from revealing the actual names of the scientists, and we have used ID instead.

The ranking methods investigated are the famous h-index [13], the C^3-index the Biplex PageRank [25], and the variation of C^3-index, namely C^4-index proposed here.

For the C^3-index we created the necessary networks as follows. Given the paper-paper citation and co-authorship paper information, a) we created undirected weighted links in Author Coauthorship layer for authors who coauthored a paper together, with the weight being the number of papers these two published together, b) we created directed unweighted links between paper using the Paper Citation information (which paper cites another paper), c) we created undirected unweighted links between authors in Author Coauthorship Layer and papers in Paper Citation Layer linking every author with every paper they published, d) we created directed weighted links in the Author Citation Layer linking every author of the citing paper to every author of the cited paper (directed to the cited ones).

The implementation of Biplex PageRank requires the existence of two layers, namely the authors's layer and papers's layer. We then created bidirectional links co-authors, between authors and their papers, directional links (from paper to paper) based on paper citations, and directional links (from author to author) based on paper citations from every author of the citing paper to every author of the cited paper. The implementation of Biplex PageRank requires to tackle also the issue that the authorship layer and the paper layer do not have exactly the same node clones. Therefore, we introduced the concept of *virtual* nodes by enforcing each layer to contain all nodes from both layers. This way, nodes that are actually present in that layer (they are included in any link) are now called *real nodes*, whereas the rest are the virtual nodes.

6.2 Visual analysis of the indices' value distribution and correlation

For the visual comparison of the algorithm results, the Standard Scores of each author were used. Then Standard Score (z) is defined as follows:

$$\zeta = \frac{x-\mu}{\sigma} \quad (15)$$

where μ is the mean value of scores and σ is the standard deviation of scores.

The results are depicted in Figures 2–6

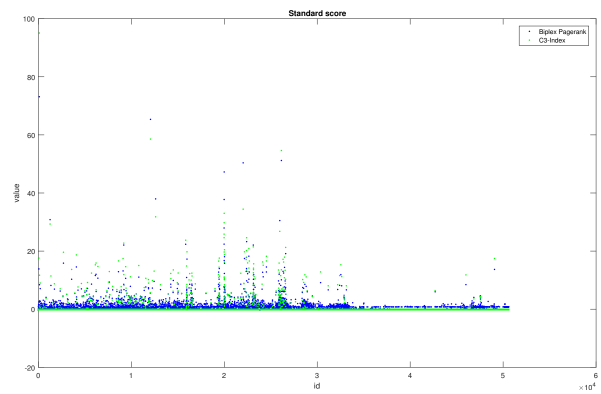

Figure 2: Biplex PageRank versus C^3-index value distribution and correlation per author ID.

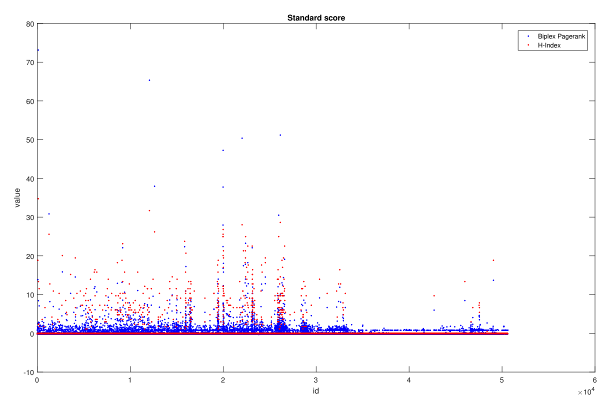

Figure 3: Biplex PageRank versus h-index value distribution and correlation per author ID.

We investigated the correlation analysis among the three existing ranking techniques. We used the Pearson correlation coefficient (PCC), which returns a measure of the linear correlation between two vectors X and Y. It takes a value in the range $[+1, -1]$, where $+1$ is total positive linear correlation, 0 is no linear correlation, and -1 is total negative linear correlation.

The results of the correlation analysis among the values of the three indices is depicted in Table 1. Clearly, C^3- and h-indices are quite highly correlated. Strong correlation between the two algorithms was expected. This is due to the strong correlation between

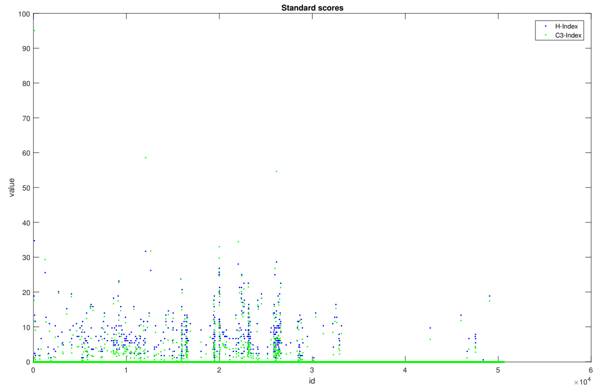

Figure 4: h-index versus C^3-index value distribution per author ID.

AAI (Author citation Index), which is one the three measures used by C^3-index (around 90%). The overall correlation drops to ≈ 75%, because of the other two measures: PCI (Paper citation Index) and AAI (Author coAuthorship Index), which return better scores for medium and low ranked researchers – that translates to better rankings for younger scientists. On the other and, Biplex PageRank's correlation to them is very low; in fact, it is quite close to no correlation at all.

	Biplex PR	h-index	C^3-index	C^4-index
Biplex PR	1	0.132232	0.101006	0.13332
h-index	0.132232	1	0.748588	0.772306
C^3-index	0.101006	0.748588	1	0.876218
C^4-index	0.13332	0.772306	0.876218	1

Table 1: Correlation coefficient among Biplex PageRank, h-index, C^3-index and C^4-index.

In Table 2 we present the top-10 scientists according to Biplex PageRank and their corresponding positions in the rankings of C^3-index and h-index. Furthermore, looking at the top-25 performing scientists according to these three algorithms we identify some commonly appearing IDs (see Table 3).

We now examine the properties of the newly proposed index, our C^4-index against its competitors.

A correlation analysis reveals that – as expected – the new index C^4-index is strongly correlated to C^3-index. However, C^4-index provides very different ranking in the top positions, and our results showed that the difference can be as high as 15% for the top-10 scientists.

In any case, the introduction of the C^4-index is just a first step towards the investigation of the wealth of information captured in multilayer networks and the exploitation of the solid mathematical theories and tools (e.g., tensor analysis) that surround them. We envision this area to be a very fertile research field.

scientist ID	C^3-index	h-index	Biplex PageRank
2037300	8	1	1
2074100	42	7	2
1719800	188	63	3
1474100	29	10	4
131520	13	2	5
2209700	14	24	6
180290	26	4	7
1028700	41	39	8
731121	132	17	9
1142000	95	66	10

Table 2: Biplex PageRank's top-10 scientists and their respective rank position in the h-index and C^3-index rankings.

scientist ID	C^3-index	h-index	Biplex PageRank
195120	6	23	12
354310	8	9	699
2037300	9	1	1
1545100	10	6	32
2209700	15	25	6
1480700	19	4	23

Table 3: Scientists (IDs) appearing at the same time in top-25 of the ranked lists of C^3-index and h-index and their respective position in Biplex PageRank's ranking.

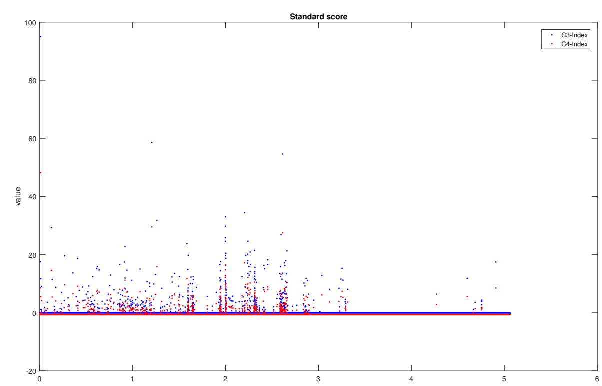

Figure 5: C^4-index versus C^3-index value distribution per author ID.

7 CONCLUSIONS

During the past decades many different measurements have been designed to evaluate the scientific impact of scholars, journals and academic institutions, which have shaped the now mature the field of scientometrics/informetrics. However, the multifacet structure, the dynamics and diffusion phenomena observed in the scientific research and subsequent outcomes, have given rise to the so-called *Science of Science* aiming at studying these mechanisms. In this article, we provide a brief survey of what is Science of Science,

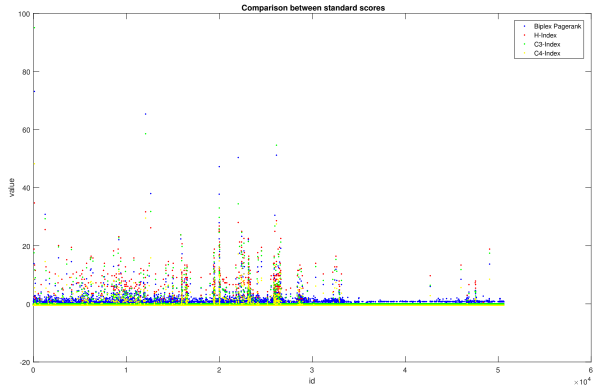

Figure 6: C^4-index versus the other competitors' value distribution per author ID.

especially with respect to scientific evaluation. Then, motivated by the recent progress in its topic related to multilayer networks, we examine the issue of scientists' ranking placed in the context of multilayer networks, by proposing a hybrid ranking index, namely C^4-index. Its aim is mainly to bring awareness about the tremendous opportunities that open in the Science of Science context rather than become a yet another index. The presentation of the new index is accompanied by a brief and comprehensive evaluation to show off its potential.

REFERENCES

[1] D. E. Acuna, S. Allesina, and K. P. Kording. Future impact: Predicting scientific success. *Nature*, 489(7415):201–202, 2012.
[2] A. L. Barabasi. *Network Science.* Cambridge University Press, 2016.
[3] A. L. Barabasi and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
[4] P. Basaras, G. Iosifidis, D. Katsaros, and L. Tassiulas. Identifying influential spreaders in complex multilayer networks: A centrality perspective. *IEEE Transactions on Network Science & Engineering*, 2018. To appear.
[5] P. Basaras, D. Katsaros, and L. Tassiulas. Detecting influential spreaders in complex, dynamic networks. *IEEE Computer magazine*, 46(4):26–31, 2013.
[6] S. Boccaletti, G. Bianconi, R. Criado, C. I. del Genio, J. Gómez-Gardeñes, M. Romance, I. Sendina-Nadal, Z. Wang, and M. Zanin. The structure and dynamics of multilayer networks. *Physics Reports*, 544:1–222, 2014.
[7] S. V. Buldyrev, R. Parshani, P. Gerald, S. H. Eugene, and H. Shlomo. Catastrophic cascade of failures in interdependent networks. *Nature*, 464:1025–1028, 2010.
[8] A. Cardillo, J. Gómez-Gardeñes, M. Zanin, M. Romance, D. Papo, F. del Pozo, and S. Boccaletti. Emergence of network features from multiplexity. *Nature Scientific Reports*, 3:1344, 2013.
[9] M. De Domenico, A. Solé-Ribalta, E. Omodei, S. Gómez, and A. Arenas. Ranking in interconnected multilayer networks reveals versatile nodes. *Nature Communications*, 6:6868, 2015.
[10] S. Fortunato, C. T. Bergstrom, K. Borner, J. A. Evans, D. Helbing, S. Milojevic, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A. L. Barabasi. Science of science. *Science*, 359(6379), 2018.
[11] M. A. Garcia-Perez. Limited validity of equations to predict the future h index. *Scientometrics*, 96(3):901–909, 2013.
[12] A. Halu, R. J. Mondragón, P. Panzarasa, and G. Bianconi. Multiplex PageRank. *PLOS One*, 8(10):e78293, 2013.
[13] J. E. Hirsch. An index to quantify an individual's scientific research output. *Proceedings of the National Academy of Sciences*, 102(46):16569–16572, 2005.
[14] L. Katz. A new status index derived from sociometric analysis. *Psychometrika*, 18(1):39–43, 53.
[15] D. Kempe, E. Tardos, and J. Kleinberg. Maximizing the spread of influence through a social network. *Theory of Computing*, 11(4):105–147, 2015.
[16] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljeros, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 6:888–893, 2010.
[17] M. Kivelä, A. Arenas, M. Barthelemy, J. P. Gleeson, Y. Moreno, and M. A. Porter. Multilayer networks. *Journal of Complex Networks*, 2:203–271, 2014.
[18] A. N. Langville and C. D. Meyer. *Google's PageRank and Beyond: The Science of Search Engine Rankings.* Princeton University Press, 2006.
[19] V. Lattora, V. Nicosia, and G. Russo. *Complex Networks: Principles, Methods and Applications.* Cambridge University Press, 2017.
[20] X. Liu, J. Bollen, M. L. Nelson, and H. van de Sompel. Co-authorship networks in the digital library research community. *Information Processing & Management*, 41(6):1462–1480, 2005.
[21] P. J. Mucha and M. A. Porter. Communities in multislice voting networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 20(4):041108, 2010.
[22] M. E. J. Newman. *Networks: An Introduction.* Oxford University Press, 2010.
[23] Z. Ning, Y. Liu, J. Zhang, and X. Wang. Rising star forecasting based on social network analysis. *IEEE Access*, 5:24229–24238, 2017.
[24] F. Papadopoulos, M. Kitsak, M. A. Serrano, M. Boguna, and D. Krioukov. Popularity versus similarity in growing networks. *Nature*, 489(7417):537–540, 2012.
[25] F. Pedroche, M. Romance, and R. Criado. A biplex approach to PageRank centrality: From classic to multiplex networks. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(6), 2016.
[26] O. Penner, R. K. Pan, A. M. Petersen, K. Kaski, and S. Fortunato. On the predictability of future impact in science. *Nature Scientific Reports*, 3(10), 2013.
[27] B. Perozzi, R. Al-Rfou, and S. Skiena. Deepwalk: Online learning of social representations. In *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, pages 701–710, 2014.
[28] D. Pradhan, P. S. Paul, U. Maheswari, S. Nandi, and T. Chakraborty. C^3-index: A PageRank based multi-faceted metric for authors' performance measurement. *Scientometrics*, 110(1):253–273, 2017.
[29] N. Shibata, Y. Kajikawa, and I. Sakata. Link prediction in citation networks. *Journal of the American Society for Information Science and Technology*, 63(1):78–85, 2012.
[30] A. Sidiropoulos, D. Katsaros, and Y. Manolopoulos. Generalized Hirsch h-index for disclosing latent facts in citation networks. *Scientometrics*, 72(2):253–280, 2007.
[31] R. Sinatra, D. Wang, P. Deville, C. Song, and A. L. Barabasi. Quantifying the evolution of individual scientific impact. *Science*, 354(6312), 2016.
[32] M. Singh, A. Jaiswal, P. Shree, A. Pal, A. Mukherjee, and P. Goyal. Understanding the impact of early citers on long-term scientific impact. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 59–68, 2017.
[33] D. Wang, C. Song, and A. L. Barabasi. Quantifying long-term scientific impact. *Science*, 342(6154):127–132, 2013.
[34] P. Wang, G. Robins, P. Pattison, and E. Lazega. Exponential random graph models for multilevel networks. *Social Networks*, 35:96–115, 2013.
[35] D. J. Watts and S. H. Strogatz. Collective dynamics of small-worldnetworks. *Nature*, 393(6684):440–442, 1998.
[36] L. Weihs and O. Etzioni. Learning to predict citation-based impact measures. In *Proceedings of the ACM/IEEE Joint Conference on Digital Libraries (JCDL)*, pages 49–58, 2017.
[37] E. Yan, Y. Ding, and C. R. Sugimoto. P-rank: An indicator measuring prestige in heterogeneous scholarly networks. *Journal of the American Society for Information Science and Technology*, 62(3):467–477, 2011.
[38] A. Zeng, Z. Shen, J. Zhou, J. Wu, Y. Fan, Y. Wang, and H. E. Stanley. The science of science: From the perspective of complex systems. *Physics Reports*, 714–715:1–73, 2017.
[39] A. Zeng and C. H. Yeung. Predicting the future trend of popularity by network diffusion. *Chaos: An Interdisciplinary Journal of Nonlinear Science*, 26(6):063102, 2016.
[40] Q.-M. Zhang, L. Lu, W.-Q. Wang, and T. Zhou. Potential theory for directed networks. *PLOS One*, 8(2):e55437, 2013.
[41] T. Zhou, L. Lu, and Y.-C. Zhang. Predicting missing links via local information. *Europhysics Journal B*, 71(4):623–630, 2009.
[42] Y.-B. Zhou, L. Lu, and M. Li. Quantifying the influence of scientists and their publications: Distinguishing between prestige and popularity. *New Journal of Physics*, 14(3):033033, 2012.