OTARIOS: OpTimizing Author Ranking with Insiders/Outsiders Subnetworks

Jorge Silva, David Aparício and Fernando Silva

Abstract Evaluating scientists based on their scientific production is a very controversial topic. Nevertheless, bibliometrics and algorithmic approaches can improve traditional peer review in numerous tasks, such as attributing research grants, deciding scientific committees, or choosing faculty promotions. Traditional bibliometrics focus on individual measures, disregarding the whole data (i.e., the whole network). Network algorithms, such as PageRank, have been used to measure node/author importance in a network. However, traditional PageRank and state-of-the-art (STOA) variations either ignore or do not combine effectively relevant information, such as the the author's productivity or the venue and year of the publication/citation. Furthermore, STOA algorithms assume that we have access to the full network which, for most real cases, is impossible. Here we put forward OTARIOS, a graph-ranking method which combines multiple publication/citation criteria to rank authors. OTARIOS divides the original network in two subnetworks, insiders and outsiders, which is an adequate representation of citation networks with missing information. We evaluate OTARIOS on a set of five real networks, each with publications in distinct areas of Computer Science. We observe that OTARIOS is $> 30\%$ more accurate than traditional PageRank and $> 20\%$ more accurate than STOA.

1 Introduction

Deciding scientific committees, research grants, or faculty promotions is still done mostly by peer review. Nevertheless, bibliometrics have been proposed that assist the peer review process [12]. Bibliometrics typically rely on the author's productivity (i.e., statistics of author's papers) and the author's impact (i.e., statistics of

Jorge Silva · David Aparício · Fernando Silva

CRACS & INESC-TEC, DCC-FCUP, Universidade do Porto, Portugal
e-mail: jmbs@inesctec.pt;{daparicio, fds}@dcc.fc.up.pt

author's citations)[1], e.g., one of the most widely used bibliometrics is the author's h-index[5], which measures the impact only of the author's most relevant works.

However, traditional bibliometrics have the drawback of only assigning impact to direct citations and ignore indirect citations. For example, if author A cites B, and B cites C, traditional bibliometrics give no additional merit to C from A's indirect citation. To tackle this problem, researchers have started developing metrics using graph algorithms on citation networks [2, 8, 3, 14, 9, 3]. These strategies are based on the PageRank algorithm [7]. One of PageRank's major ideias is that not all nodes are equal, i.e., it is good to be referenced by any page but it is better to be referenced by important pages. This idea also applies to citation networks, where it is important to be cited by important authors and not just by any author. State of the art ranking algorithms for citations networks adapt PageRank and introduce modifications to favour different types of authors (e.g., authors cited in important venues, or authors cited more recently).

We find that state of the art is lacking in two aspects. First, these methods do not adequately combine publication attributes (e.g., author's productivity, the venues prestige of where the usually publishes, and how recent his papers are) with citation attributes (e.g., the prestige of the venue he is being cited from, how recent his citations are). Second, these methods assume that the full network is known and the algorithm does not distinguish between *fully explored nodes* and *partially explored nodes*.

Here we propose a graph-ranking algorithm to rank scientists/authors, named OTARIOS (**O**p**T**imizing **A**uthor **R**anking with **I**nsiders/**O**utsiders **S**ubnetworks). OTARIOS handles the first problem by efficiently combining different publication/citation attributes in a multi-edge weighted network (instead of a simple unweighted network used by state of the art methods). OTARIOS is also a flexible algorithm in the sense that publication/citation attributes can be personalised to fit what the user wants the researchers to be ranked by (e.g., value venue prestige highly or lowly). OTARIOS handles the second problem by dividing the citation network in two subnetworks, *insiders* and *outsiders*. Then, only insiders are ranked (since we have their full information) while outsiders contribute to the ranks of insiders, not being themselves ranked. Our results on five networks belonging to different areas of Computer Science show that OTARIOS is $> 20\%$ more accurate than state of the art methods.

The paper is organised as follows. Section 2 describes terminology that is used throughout the work, as well as an overview of state of the art methods. Section 3 describes OTARIOS and our methodology. Section 4 presents the performance of OTARIOS against state of the art methods on a set of five networks. Finally, Section 5 presents our main conclusions and gives some directions for future work.

2 Preliminaries

2.1 Terminology

Recency of a paper

$$\delta(P_j) = \left(\max_{P_{j'} \in \mathcal{P}} y(P_{j'})\right) - y(P_j) \quad (1)$$

Recency of an author

$$\delta(A_i) = \min_{P_j \in \mathcal{P}_{A_i}} \delta(P_j) \quad (2)$$

Venue prestige

$$\lambda(V_k, y) = \frac{c(V_k, y)}{\sum_{x=1}^{3} p(V_k, y - x)} \quad (3)$$

Cited individuality

$$w(A_{i'} \to A_i, P_j) = \frac{1}{|\mathcal{A}_{P_j}|}, A_i \in \mathcal{A}_{P_j} \quad (4)$$

Citation recency

$$a(A_{i'} \to A_i, P_j) = e^{\frac{-\delta(P_j)}{\tau}}, A_{i'} \in \mathcal{A}_{P_j} \quad (5)$$

Citation prestige

$$v(A_{i'} \to A_i, P_j) = v(P_j), A_{i'} \in \mathcal{A}_{P_j} \quad (6)$$

For consistency, we denote sets by calligraphic letters (e.g., $\mathcal{S}$), elements of those sets (i.e., entities) by capital letters with an index (e.g., S_i), attributes of entities (e.g., year, impact factor) as functions named in lower-case alphabetic or greek letters (e.g., $a(S_i)$ or $\alpha(S_i)$) and constants as sole greek letters (e.g., τ). Cardinality of a given set S is denoted by $|\mathcal{S}|$. We address the following problem.

Problem 1. Given a set of papers $\mathcal{P}$ published in a set of venues $\mathcal{V}$ by a set of authors $\mathcal{A}$, who are the n top-ranked authors?

A paper $P_j \in \mathcal{P}$ is co-authored by authors $\mathcal{A}_{P_j} \subseteq \mathcal{A}$. Likewise, an author $A_i \in \mathcal{A}$ is (one of) the author(s) of papers $\mathcal{P}_{A_i} \subseteq \mathcal{P}$. In paper-level networks, graph $G = \{\mathcal{N}, \mathcal{E}\}$ comprises a set $\mathcal{N}$ of nodes that represent papers and a set $\mathcal{E}$ of edges that represent paper citations, written as $P_{j'} \to P_j$. In author-level networks, nodes represent authors and edges represent citations between authors, written as $A_{i'} \to A_i$.

Regarding node attributes, papers have publication metadata which we use as attributes, namely the year, venue prestige, and the number of references, represented by $y(P_j)$, $v(P_j)$ and $r_{out}(P_j)$, respectively. The *recency* of a paper, represented by $\delta(P_j)$, is given by Equation 1. Similarly, the *recency* of an author, represented by $\delta(A_i)$, is simply the recency of his most recent paper (Equation 2). The venue prestige of a paper P_j depends on the venue $V_k \in \mathcal{V}$ where it was published and the year when it was published, represented by $v(P_j) = \lambda(V_k, y(P_j))$. We estimate venue prestige with *CiteScore*, a widely used metric[1] (Equation 3), where $p(V_k, y)$ is the number of papers published in V_k in year y and $c(V_k, y)$ is the number of citations that all papers published in V_k in year y received. We are thus giving higher prestige to venues that have many citations per paper.

Regarding edges, in paper-level networks edges are traditionally unweighted and simple, i.e., two papers are connected by a single edge with weight equal to 1 [6, 3].

In author-level networks, edges are weighted and multiple, i.e., two authors are connected by multiple edges with different weights. These multiple edges concern different edge attributes that depend on the publication P_j where author $A_{i'}$ cites author A_i. The recency of an edge, represented by $a(A_{i'} \rightarrow A_i, P_j)$, gives more importance to recent citations (Equation 5). As discussed in the NewRank paper which originally proposes this concept for author ranking algorithms [3], we set the decay factor $\tau = 4$. The venue prestige of an edge, represented by $v(A_{i'} \rightarrow A_i, P_j)$, gives more importance to citations in important venues (Equation 6). Finally, the individuality of an edge, represented by $w(A_{i'} \rightarrow A_i, P_j)$, gives more importance to citations received in papers that author A_i has few (or no) co-authors (Equation 4). Thus, $w(A_{i'} \rightarrow A_i, P_j)$, unlike $a(A_{i'} \rightarrow A_i, P_j)$ and $v(A_{i'} \rightarrow A_i, P_j)$, depends on the cited author A_i and not on the citing author $A_{i'}$. The author's attribute total out-edge weight is obtained by summing all of its out-edges, e.g., for citation recency, $a_{out}(A_i) = \sum_{(A_i \rightarrow A_{i'}, P_j)} a(A_i \rightarrow A_{i'}, P_j)$ – w_{out} and v_{out} are obtained in the same way.

2.2 State of the art

Measuring the scientific impact of institutions, journals, or authors is an important task in the peer review process. Here we focus on measuring the impact of authors, i.e., author ranking. Traditional bibliometrics, such as the widely used h-index [5], evaluate an author's impact simply by the number of citations of his most relevant papers. Other metrics, such as CountRank[4], measure scientific impact as a combination of the author's productivity (i.e., the number of papers that he publishes) and the author's popularity (i.e., the number of citations that he receives). However, these metrics fail to capture the nature of scientific development since they disregard the fact that a new discovery is not due solely to previous work directly referenced. Graph-based metrics, on the other hand, correctly spread the credit to previous works that paved the way [13].

There are two groups of graph-ranking methods: paper-level and author-level [13]. On one hand, paper-level ranking uses the papers' citation network to diffuse scientific credit to cited papers, and then author credit is derived from the credit of his papers [6, 3]. On the other hand, author-level ranking uses the authors' citation network to diffuse scientific credit to cited authors, thus the authors' credit is directly obtained [8, 2, 14]. While these two methods give origin to different networks (Figure 1), PageRank-like [7] methods are typically used to measure node importance in both.

The score initialisation step creates a vector R that defines an initial score for every node using *a priori* information. In the simplest case, every node (i.e., paper or author) is considered equally important, thus an uniform distribution is used (i.e., $R[i] = \frac{1}{N}$, where i is a node and N is the number of nodes in the network) [2, 9]. Approaches based on paper citation networks typically assign higher initial scores to more recent papers [3] or favour a combination of recent papers and papers pub-

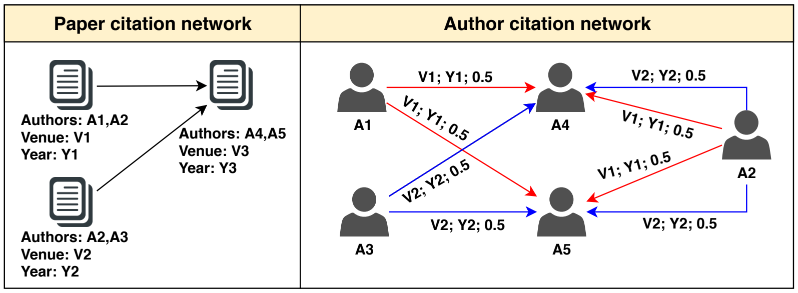

Fig. 1 Comparison of paper-level and author-level networks.

lished in venues with higher impact factor [6]. Approaches based on author citation networks typically assign higher initial scores to authors that publish many papers [14] or favour authors that publish many papers but with few co-authors [8].

The score diffusion step updates the node scores taking into consideration network dynamics. Score diffusion is an iterative process which computes three addends: random restart, dangling nodes, and score term. Random restart (RR) evaluates how likely it is to reach a certain node by moving randomly in the network. Graph-ranking algorithms define a value q as the random restart probability, and q is multiplied by the node's initial score R (thus, nodes with higher initialisation receive higher random restart score). Dangling nodes (DN) is a process where the score of nodes that do not have any out-links is split by all other nodes. This is performed to avoid having nodes that do not disseminate their credit. Like random restart, this division takes into consideration the initialisation vector R, thus nodes initialised with higher values give more credit to other nodes. Score term (ST) updates the score of a node i, according to the score of his in-links (i.e., nodes citing i). There are several strategies to distribute this score, and the simplest ones simply take into consideration the weight defined in the citation network (i.e., distribute the score evenly by co-authors of the cited publication, e.g., if the papers has two authors, the score is divided by the two authors, if it has three authors, the score is divided by the three authors, thus, in the case of three authors, each author receives less credit than in the case with just two authors) [8, 2, 14]. SCEAS [9] adds a constant value b to the every score received by nodes and divides the total score received by another constant a in order to make the algorithm converge faster. YetRank and NewRank [6, 3] take into consideration the vector R in the score distribution (i.e., if a paper cites a paper P_a from 2015 and another paper P_b from 2010, P_a receives a bigger chunk of the score. In case of the YetRank, the distribution of score also takes into consideration the impact factor of the venues where P_a and P_b where published, favouring papers published in venues higher prestige.

Table 1 summarises state of the art graph-ranking methods and their differences.

One drawback of current graph-ranking approaches is that they assume that the complete citation network is known. However, in real-world cases, it is not possible to obtain a complete network. Let us consider an author-level citation network: to rank a set of authors $\mathcal{A}$, we extract all authors $\mathcal{B}$ that cite any $A_i \in \mathcal{A}$. Then, we

<table>
<tr><th></th><th>Method</th><th>Initialisation: $R(N_i)$</th><th>Score term: $ST(N_i)$</th></tr>
<tr><td rowspan="4">Author-level</td><td>RLPR [2]</td><td>$\frac{1}{|\mathcal{A}|}$</td><td rowspan="3">$(1-q)\sum_{(A_{i'} \to A_i, P_j)} \frac{S(A_{i'}) \times w(A_{i'} \to A_i, P_j)}{w_{out}(A_{i'})}$</td></tr>
<tr><td>SARA [8]</td><td>$\frac{\sum_{(P_j \in \mathcal{P}_{A_i})} \frac{1}{|\mathcal{A}_{P_j}|}}{\sum_{(A_{i'} \in \mathcal{A})} \sum_{(P_j \in \mathcal{P}_{A_{i'}})} \frac{1}{|\mathcal{A}_{P_j}|}}$</td></tr>
<tr><td>ALEF [14]</td><td>$\frac{|\mathcal{P}_{A_i}|}{|\mathcal{P}|}$</td></tr>
<tr><td>SCEAS [9]</td><td>$\frac{1}{|\mathcal{A}|}$</td><td>$\frac{(1-q)}{a}\sum_{(A_{i'} \to A_i, P_j)} \frac{(S(A_{i'})+b) \times w(A_{i'} \to A_i, P_j)}{w_{out}(A_{i'})}$</td></tr>
<tr><td rowspan="2">Paper-level</td><td>YetRank [6]</td><td>$v(P_i) \times \frac{e^{\frac{-\delta(P_i)}{\tau}}}{\tau}$</td><td rowspan="2">$(1-q)\sum_{(P_{i'} \to P_i)} \frac{S(P_{i'}) \times R(P_i)}{r_{out}(P_{i'})}$</td></tr>
<tr><td>NewRank [3]</td><td>$e^{\frac{-\delta(P_i)}{\tau}}$</td></tr>
</table>

Table 1 Comparison of state of the art methods. N_i represents a node in the network, i.e., $N_i = A_i$ in author-level networks, and $N_i = P_i$ in paper-level networks. Score diffusion $S(N_i)$ is equal to $ST(N_i) + RR(N_i) + DN(N_i)$. For all methods, $RR(N_i) = q \times R(N_i)$ and $DN(N_i) = (1-q) \times R(N_i)$, thus we omit them from the table.

need to extract all authors $\mathcal{C}$ that cite any $B_i \in \mathcal{B}$, to correctly determine the scores of all $A_i \in \mathcal{A}$, i.e., $C_i \in \mathcal{C}$ does not cite $A_i \in \mathcal{A}$ directly but he cites some B_i that cites A_i, thus C_i cites A_i indirectly. Ideally, this should be performed until the complete set of citations with seed $\mathcal{A}$ is obtained. Due to memory and time constraints, current algorithms only obtain a sample of the full network but disregard that some information was lost. As a result, current state of the art graph-ranking algorithms are estimating scientific rankings based on incorrect information, i.e., authors in the periphery are not being adequately taken into account since their citations are not in the network. Although there is no ideal solution for this problem, one can be more careful in estimating the rank of nodes in the periphery.

In this paper we put forward OTARIOS, a novel graph-ranking algorithm for authors. OTARIOS uses the concept of *outsiders* to estimate the rank of nodes/authors in the periphery and, thus, does not require the complete citation network. Furthermore, OTARIOS efficiently combines node properties (i.e., statistics about the authors' publicatitons) with edge properties (i.e., statistics about the author's citations), information that state of the art methods mostly disregard (Table 2).

Method	Publications: Volume	Publications: Recency	Publications: Venues	Citations: Individuality	Citations: Recency	Citations: Venues	Limited Info.
RLPR				✓			
SARA	✓			✓			
ALEF	✓*			✓			
SCEAS				✓			
YetRank		✓	✓	✓			
NewRank		✓		✓			
OTARIOS	✓	✓	✓	✓	✓	✓	✓

Table 2 Comparison of state of the art methods with OTARIOS. OTARIOS tries to combine all criteria efficiently and is also the only method that adequately deals with networks with limited information by using insiders/outsiders subnetworks.
*ALEF gives higher score to authors with many publications but ignores the number of authors in the publications.

3 Methodology

3.1 Problem Description

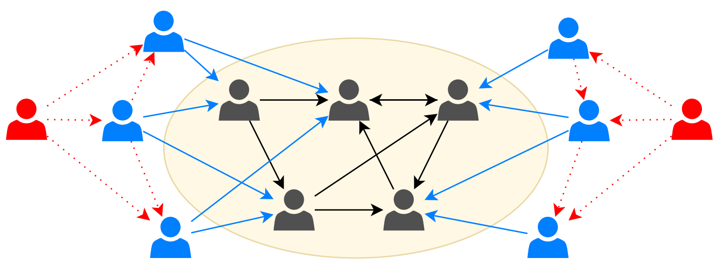

Fig. 2 Example of insiders and outsiders subnetworks. Insiders are nodes/authors coloured in black and outsiders are coloured in blue. Note that no links between outsiders exists (dashed red lines). Furthermore, no information exists of outsiders that do not cite any insiders (coloured in red).

We propose a new author-level graph-ranking methodology. We assume that we want to rank a specific set of authors $\mathscr{I}$ (e.g., authors that publish in certain conferences, or in certain countries). First, we obtain all citations between all authors $I_i, I_{i'} \in \mathscr{I}$ (i.e., we obtain a complete citation network for $\mathscr{I}$). Second, for each author I_i, we obtain all of his received citations coming from any author $O_i \notin \mathscr{I}$. The process stops here, i.e., we do not obtain all received citations for authors $O_i \in \mathscr{O}$. Doing so repeatedly is very costly computationally and unfeasible in practice because the number of authors added at each step grows very rapidly.

We thus divide the citation network into two groups of nodes: *insiders* ($\mathscr{I}$) and *outsiders* ($\mathscr{O}$) (Figure 2). Thus, the whole set of authors $\mathscr{A} = \{\mathscr{I}, \mathscr{O}\}$. Moreover, $\mathscr{I} \cap \mathscr{O} = \emptyset$ since no outsider can also be an insider, and vice-versa. With respect to edges in the network, there are edges between insiders or from an outsider to an insider, but no edges exist from insiders to outsiders nor between outsiders (i.e.,$\neg\exists(A_{i'} \rightarrow A_i \in \mathcal{E}) : A_i \in \mathcal{O}$). The set of edges connecting insiders is represented as $\mathcal{E}_{\mathcal{I}}$ and the set of edges connecting outsiders to insiders is represented as $\mathcal{E}_{\mathcal{O}}$, thus, $\mathcal{E} = \{\mathcal{E}_{\mathcal{I}}, \mathcal{E}_{\mathcal{O}}\}$.

We aim to estimate the prestige of insiders (i.e., obtain a rank). We do not rank outsiders, instead we use them to increase the accuracy of ranks calculated for insiders. Information about the outsiders is limited (i.e., we do not have information about who cites them) but we assume that some outsiders are more important than others. We estimate outsiders' prestige (λ) before insiders rank initialisation. We use the outsiders' history of publications, giving higher prestige to authors with many citations ($c(A_i)$) in few publications ($p(A_i)$) (Equation 7). Our objective is to increase the initial rank of insiders that are cited by outsiders with high prestige.

$$\lambda(A_i) = \frac{c(A_i)}{p(A_i)} \tag{7}$$

3.2 OTARIOS

OTARIOS is a graph-based algorithm for author-level citation networks. Its aim is to rank authors based on their publication and citation history. OTARIOS uses the notion of insider/outsider subnetworks to adequately estimate authors scores in a network with limitation information. Furthermore, OTARIOS is a flexible algorithm that analyses which set of publication/citation attributes lead to better rankings.

On the first step, OTARIOS computes an initial score for each author, represented by $R(A_i)$. OTARIOS calculates $R(A_i)$ by taking into account multiple criteria that favour different author characteristics (Table 3). We divide the criteria into two categories: productivity and outsiders influence. Productivity gives credit to authors considering the value of their previously published work, while outsider influence gives credit to authors considering the value of the outsiders that cite them. Regarding productivity, OTARIOS takes three factors into account: volume (i.e., the amount of papers the author published in function of the number of co-authors), recency (i.e., the number of years since the author's last publication) and venues (i.e., the prestige of the venues where the author has published). Regarding outsiders influence, OTARIOS takes another three factors into account: individuality (i.e., exclusiveness of citations received), recency (i.e., how recent the author's citations are) and venues (i.e., the prestige of the venues where the author's citations come from). We compute the author's initial score $R(A_i)$ as the sum of the two products of the factors in each group (i.e., productivity ($volume \times recency \times venues$) + outsiders influence ($individuality \times recency \times venues$)).

On the second step, OTARIOS computes the rank scores in an iterative process. At this step outsiders are not considered part of the network. We remove them from the network since their presence degrades the score diffusion step, as we discuss and evaluate in the results section. In each iteration, OTARIOS updates an author's score $S(A_i)$ as $ST(A_i) + RR(A_i) + DN(A_i)$. Like state of the art methods, we com-

	Criteria	Initialisation: $R(A_i)$	Description
Productivity	Volume (P)	$\frac{\sum_{(P_j \in \mathcal{P}_{A_i})} \frac{1}{\lvert \mathcal{A}_{P_j} \rvert}}{\sum_{(A_{i'} \in \mathcal{A})} \sum_{(P_j \in \mathcal{P}_{A_{i'}})} \frac{1}{\lvert \mathcal{A}_{P_j} \rvert}}$	Favours publishing many papers with few co-authors.
	Recency (A)	$e^{\frac{-\delta(A_i)}{\tau}}$	Favours publishing recently.
	Venues (V)	$\Big(\sum_{(P_j \in \mathcal{P}_{A_i})} v(P_j)\Big) \times \lvert \mathcal{P}_{A_i} \rvert^{-1}$	Favours publishing in prestigious venues.
Outsiders Influence	Individuality (W)	$\sum_{(A_{i'} \to A_i, P_j)} \frac{\lambda(A_{i'}) \times w(A_{i'} \to A_i, P_j)}{w_{out}(A_{i'})}, A_{i'} \in \mathcal{O}$	Favours being cited by outsiders that cite few authors.
	Recency (A)	$\sum_{(A_{i'} \to A_i, P_j)} \frac{\lambda(A_{i'}) \times a(A_{i'} \to A_i, P_j)}{a_{out}(A_{i'})}, A_{i'} \in \mathcal{O}$	Favours being cited by outsiders more recently.
	Venues (V)	$\sum_{(A_{i'} \to A_i, P_j)} \frac{\lambda(A_{i'}) \times v(A_{i'} \to A_i, P_j)}{v_{out}(A_{i'})}, A_{i'} \in \mathcal{O}$	Favours being cited by outsiders in prestigious venues.

Table 3 List of criteria used for OTARIOS' author rank initialisation: $R(A_i)$. OTARIOS considers both the authors' productivity and the direct influence of outsiders on the authors. We use different combinations of these criteria to create different variants, e.g., OTARIOS($PV+V$), or simply $PV+V$ for brevity, uses volume (P) and venue prestige (V) to measure author productivity, and uses venue prestige (V) to measure the direct influence of outsiders.

pute $RR(A_i)$ and $DN(A_i)$ in function of the initial rank of each author, and compute $ST(A_i)$ in function of the author's citations coming from other authors (Table 1). Similar to how outsiders influence is calculated in the rank initialisation step, OTARIOS considers three different criteria to assess insiders influence (i.e., score term): individuality, recency and venues (Table 4). Like for rank initialisation, the $ST(A_i)$ at each iteration is the product of every criteria.

Here we do not assume that every criteria is equally important and that they should all be used for author ranking. The criteria's importance depends greatly on the dataset. For instance, venue prestige might be very important to rank some communities (i.e., top authors publish in top conferences of that scientific area, e.g., machine learning) but irrelevant in some other community because we are studying a specific conference (i.e., all authors published in the same venue, e.g., KDD). OTARIOS is not a single static algorithm, instead it is a flexible algorithm that uses different user-defined criteria for author ranking. For example, for a certain application, we may want to rank authors taking into account productivity with recency, outsiders influence with venues and individuality, and score term with venues. While for another application we may want to give credit only based on citations (i.e., disregarding author publications) and use a variant that only considers outsider influence using recency and score term using venues [1].

[1] Note that we define variants using notation $APV+AVW+AVW$, where the addends define the criteria used at each group. The first for productivity, the second for outsiders influence and the last

Criteria	Score term: $ST(A_i)$	Description
Individuality (W)	$\sum_{(A_{i'} \to A_i, P_j)} \frac{S(A_{i'}) \times w(A_{i'} \to A_i, P_j)}{w_{out}(A_{i'})}, A_{i'} \in \mathcal{I}$	Favours being cited by insiders that cite few authors.
Recency (A)	$\sum_{(A_{i'} \to A_i, P_j)} \frac{S(A_{i'}) \times a(A_{i'} \to A_i, P_j)}{a_{out}(A_{i'})}, A_{i'} \in \mathcal{I}$	Favours being cited by insiders more recently.
Venues (V)	$\sum_{(A_{i'} \to A_i, P_j)} \frac{S(A_{i'}) \times v(A_{i'} \to A_i, P_j)}{v_{out}(A_{i'})}, A_{i'} \in \mathcal{I}$	Favours being cited by insiders in prestigious venues.

Table 4 List of criteria used for OTARIOS' author score term calculation: $ST(A_i)$. Combined with author initialisation (Table 3), we create different variants, e.g., PV+V+A combines initialisation PV+V with score term calculation A, i.e., using citation recency. Like state of the art methods, $S(A_i) = ST(A_i) + RR(A_i) + DN(A_i)$. $RR(A_i)$ and $DN(A_i)$ are calculated as described in Table 1.

4 Results

We compare OTARIOS against state of the art methods over five created networks, each consisting of top-tier conferences in computer science (Table 5). For each network, we created a ground truth ranking using the best paper award information for every conference [2]. Each awarded paper has a unit of prestige which is equally divided by its authors. In the end, the final author-ranking consists of the sum of the prestige obtained from the awards. As a result, we are assuming that authors that have won more awards with fewer co-authors should be ranked higher in our experiments.

Network	Conferences	Nodes $\lvert\mathcal{I}\rvert$	Nodes $\lvert\mathcal{O}\rvert$	Edges $\lvert\mathcal{E}_{\mathcal{I}}\rvert$	Edges $\lvert\mathcal{E}_{\mathcal{O}}\rvert$
CM	AAAI, IJCAI, ICML, ACL, ICCV, CVPR	35.6k	224.9k	4.6M	4.9M
TC	FOCS, SODA, STOC	5.0k	82.4k	0.5M	0.8M
NET	INFOCOM, NSDI, SIGCOMM, MOBICOM, SIGMETRICS	15.2k	138.8k	2.1M	3.7M
IS	KDD, CIKM, PODS, SIGMOD, VLDB, WWW, SIGIR	282.7k	190.9k	4.0M	5.1M
SE	PLDI, FSE, ICSE, OSDI, SOSP	10.8k	99.9k	1.0M	2.1M

Table 5 Set of networks used for experimental evaluation. Data was taken from [11, 10]. The full DBLP dataset contains over 3M publications from 1936 to 2018. Each network contains publications from only a set of conferences, e.g., networks TC contains publications from FOCS, SODA and STOC. For each network we show the number of insider and outsider nodes, $|\mathcal{I}|$ and $|\mathcal{O}|$ respectively, and the number of insider and outsider edges, $|\mathcal{E}_{\mathcal{I}}|$ and $|\mathcal{E}_{\mathcal{O}}|$ respectively.

In our experiment, methods that produce rankings more similar to the ground truth one (obtained by human judgement) are better. For the purpose, for every net-

for score term. For the first example on the text, the variant nomenclature is $A+VW+V$, while for the second it is $\emptyset + A + V$.

[2] Awards information obtained from: `https://jeffhuang.com/best_paper_awards.html`

work and method, we compare the method's predicted ranking with the network's ground truth using two commonly used ranking quality measures: Normalized Discounted Cumulative Gain (NDCG) and Mean Reciprocal Rank (MRR). NDCG penalises predicted rankings that rank less prestigious authors above more prestigious ones [3], while MRR is the mean predicted ranking position for a set of authors. Commonly, both measures are estimated considering only the top @n authors. In the case of NDCG, it estimates the number of incorrect ranking placements at the top @n authors from the predicted ranking. The NDCG values range between 0 and 1, with 1 indicating a perfect ranking (i.e. all the authors have an higher or equal ground truth prestige than all the others above them). On the other hand, MRR estimates the mean predicted rank for the top @n authors from the ground truth, thus lower MRR values imply a better predicted ranking since the top authors are on average closer to the top.

For a more detailed analysis of our tests, we estimated both metrics considering different sizes of top authors ($@5, @10, @20, @50, @100$). Table 6 shows the results obtained for some of the OTARIOS variants in the network NET. The results demonstrate the process of tuning the OTARIOS criteria in order to find the variant that yields the best results. For the NET network our best variant (with an average of 0.330 NDCG and an average of 606 MRR) used productivity with recency and volume, outsiders influence with recency, and score term with recency and individuality.

OTARIOS	**NDCG**						**MRR**					
variant	@5	@10	@20	@50	@100	Mean	@5	@10	@20	@50	@100	Mean
∅ + A + ∅	0.269	0.233	0.207	0.186	0.174	0.214	443	1125	903	1526	2066	1213
∅ + V + ∅	0.269	0.233	0.207	0.186	0.185	0.216	412	1108	916	1522	2096	1211
∅ + AV + ∅	0.269	0.233	0.207	0.186	0.177	0.215	419	1109	902	1511	2074	1203
AP + A + ∅	0.288	0.246	0.259	0.218	0.241	0.250	350	500	440	1121	1502	783
AP + V + ∅	0.288	0.246	0.258	0.218	0.239	0.250	344	489	439	1134	1527	787
AP + AV + ∅	0.288	0.246	0.259	0.218	0.240	0.250	345	494	439	1143	1523	789
AP + A + A	0.380	0.297	0.283	0.282	0.280	0.304	385	647	472	1111	1416	806
AP + A + V	0.350	0.261	0.217	0.203	0.203	0.247	255	726	617	1251	1615	893
AP + A + AV	**0.407**	0.345	0.291	0.291	0.274	0.322	242	614	473	1116	1455	780
AP + A + AW	0.381	**0.369**	**0.313**	**0.302**	**0.288**	**0.330**	**219**	**386**	**328**	**879**	**1219**	**606**

Table 6 Comparison of OTARIOS variants on network NET (from Table 5). For each OTARIOS variant, we measure its ranking's NDCG and MRR for the top-5, top-10, top-20, top-50 and top-100 authors, as well as the metric's mean value. In bold we highlight the highest score for each metric. The best OTARIOS variant is coloured in blue.

[3] The prestige of an author is determined from the ground truth.

existem outro tipo de analises mais detalhadas que podemos fazer mas que nao ha espaco, Por exemplo analisar o melhor metodo por conferencia so olhando para o SOA e focar a importancia de em algumas redes escolher como criterio a idade, noutras venues etc... Depois defender que no meio desta incerteza toda, os nossos metodos foram mais robustos etc...

We compared OTARIOS with the state of the art algorithms discussed in Section 2.2 and a baseline method. Commonly, ranking algorithms compare themselves against the h-index metric. However, in our case, we are using information from the DBLP dataset which does not provide this information. Moreover, we have more than 300k different authors meaning that it is impossible to collect this information for all the authors. As a result we used CountRank (CR) as baseline. This method is a simpler version of the h-index that counts the citations received by each author. We created three CR variants: uniform, individuality and position. For each citation received, uniform assigns the same merit to all of the authors in publication (1), individuality equally divides the merit for all the authors ($\frac{1}{|\mathcal{A}|}$), and position gives more credit to authors which name appears first in the publication (first author: 1, second author: $\frac{1}{2}$, third author: $\frac{1}{3}, ...$). Table 7 shows the results obtained for all the state of the art methods and 5 OTARIOS variants over all networks. For each network, we obtained the mean of both metrics ($@5, @10, @20, @50 and @100$)). Furthermore, we estimated the mean metric value obtained over all networks. The results demonstrate that *SCEAS* is the best method state of the art method, obtaining a maximum NDCG mean of 0.208 and a best MRR mean of 691. The $CR_{position}$ method presented the worst mean for the NDCG (0.154), while *NewRank* obtained the worst (by a high margin) MRR mean of 4091. Another important aspect to highlight is that $CR_{individuality}$ despite being a baseline strategy, produced the second best results for NDCG and MRR.

mais uma vez, outras analises que podiamos fazer, mas para as quais nao temos espao, era ver que metodos que usam recency parecem sem melhores do que aqueles que usam venues para este problema e etc...

With respect to OTARIOS variants, we tested 53 variants and in total, 21 produced better *MRR* and *NDCG* results than the best state of the art method (*SCEAS*). Our variants were able to obtain a best score of 0.246 NDCG mean and 567 MRR mean. Our most robust variant in both metrics and the one that we consider our best method (*AP+A+AW*), uses productivity with criteria recency and volume, outsiders influence with recency and the score term with recency and individuality. This variant obtained a mean NDCG of 0.245 and a mean MRR of 570. We compared the gain of this variant with respect to state of the art methods, using equations 8 and 9. Compared to *RLPR*, a traditional PageRank algorithm applied to author citation networks, we obtained a gain of 28% for the NDCG mean and 27% for the MRR mean. With respect to the best state of the art method, we improved the NDCG mean value by 18% and the MRR mean by 21%.

$$G_{NDGC} = \frac{\text{OTARIOS}_{<NDGC>} - \text{STOA}_{<NDGC>}}{min(\text{OTARIOS}_{<NDGC>}, \text{STOA}_{<NDGC>})} \quad (8)$$

	Method	NDGC CM	TC	NET	IS	SE	Mean	MRR CM	TC	NET	IS	SE	Mean
State of the art	$CR_{position}$	0.097	0.049	0.189	0.176	0.261	0.154	1427	463	1009	892	324	823
	YetRank	0.128	0.028	0.206	0.157	0.271	0.158	2083	673	1047	846	491	1028
	ALEF	0.152	0.020	0.182	0.129	0.323	0.161	1260	561	670	803	310	721
	$CR_{uniform}$	0.138	0.045	0.278	0.189	0.222	0.174	1659	516	1066	1067	387	939
	RLPR	0.180	0.032	0.231	0.176	0.338	0.191	1203	508	817	720	356	721
	SARA	**0.193**	0.035	0.232	0.156	0.354	0.194	1122	461	738	**668**	303	658
	NewRank	0.115	0.004	0.297	**0.319**	0.266	0.200	5057	3112	3597	6637	2050	4091
	$CR_{individuality}$	0.129	0.043	0.247	0.211	0.372	0.200	1171	438	878	744	289	704
	SCEAS	0.143	0.035	0.275	0.255	0.335	0.208	1154	493	776	752	279	691
OTARIOS	∅ + AVW + AW	0.143	0.081	0.323	0.213	0.315	0.215	1161	324	664	707	289	629
	∅ + V + AW	0.148	0.080	0.321	0.214	0.314	0.215	1169	325	671	709	294	634
	AP + VW + AW	0.150	**0.087**	0.330	0.268	**0.383**	0.244	**1070**	273	**604**	680	**207**	**567**
	AV + VW + AW	0.143	0.085	**0.356**	0.264	**0.383**	**0.246**	1333	285	618	676	215	626
	AP + A + AW	0.152	**0.087**	0.330	0.273	**0.383**	0.245 (+18%)	1079	**272**	606	688	**207**	570 (+21%)

Table 7 Comparison of state of the art (STOA) methods against OTARIOS over all networks. The value of each cell is the metric's mean value for that network (e.g., the mean NDCG and MRR of AP+A+AW for network NET is highlighted in Table 6). In bold we highlight the highest score for each metric. The best STOA method (i.e., SCEAS) is colored in red and the best OTARIOS variant is colored in blue. Inside parentheses we show the gain of OTARIOS versus SCEAS, i.e., G_{NDGC} and G_{MRR}, respectively.

$$G_{MRR} = \frac{\text{STOA}_{<MRR>} - \text{OTARIOS}_{<MRR>}}{min(\text{OTARIOS}_{<MRR>}, \text{STOA}_{<MRR>})} \tag{9}$$

4.1 More is not always better

In our previous experiments, for the state of the art methods results, we only considered the author citation networks for the insiders authors (i.e. outsiders authors were not part of the network). However, for the OTARIOS, since we require outsiders for some of our criteria, we used a network consisting of insiders and outsiders. In order to demonstrate that we were not unfairly comparing our variants with other methods with less information, we tested their algorithms using the complete network and compared the results with the ones obtained only on the insiders network. Table 8 shows the results of this comparison [4]. The results demonstrate that on average, the state of the art methods obtained a negative gain of -17% for NDCG and -25% for MRR when using the complete network. The *NewRank* and *SCEAS* methods were the ones that presented the worst gains (-54% and -30% on NDCG, and -63% amd -37% on MRR). These methods were among the top state of the art methods when considering only the insiders network, as a result the complete network had a higher impact when compared to other methods that were already not too good. The only method that presented an overall positive gain was the *YetRank* one on the NDCG metric. However, this was a very small gain, and it came as no

[4] Gains estimated using equations 8 and 9

Method	NDCG						MRR					
	CM	TC	NET	IS	SE	Mean	CM	TC	NET	IS	SE	Mean
SCEAS	**0.144**	**0.036**	**0.275**	**0.255**	**0.335**	**0.209**	**1154**	**493**	**776**	**753**	**279**	**691**
Fullnet SCEAS	0.106	0.024	0.224	0.198	0.250	0.160	1517	845	929	999	433	945
Gain	-36%	-51%	-23%	-29%	-34%	-30%	-31%	-71%	-20%	-33%	-55%	-37%
SARA	**0.194**	**0.036**	**0.232**	0.157	**0.355**	**0.195**	**1123**	**461**	**739**	**668**	**303**	**659**
Fullnet SARA	0.181	0.030	0.227	**0.177**	0.300	0.183	1146	602	885	719	408	752
Gain	-7%	-20%	-2%	+13%	-18%	-6%	-2%	-31%	-20%	-8%	-35%	-14%
RLPR	**0.181**	**0.032**	**0.231**	**0.177**	**0.338**	**0.192**	**1203**	**508**	**817**	**721**	**357**	**721**
Fullnet RLPR	0.174	0.027	0.227	0.162	0.276	0.173	1274	728	864	757	436	812
Gain	-4%	-18%	-2%	-9%	-22%	-11%	-6%	-43%	-6%	-5%	-22%	-13%
ALEF	**0.152**	0.021	0.183	0.130	**0.323**	**0.162**	**1261**	**561**	**670**	804	**310**	**721**
Fullnet ALEF	0.125	**0.024**	**0.203**	**0.151**	0.299	0.160	1373	608	930	**735**	432	816
Gain	-22%	+13%	+11%	+16%	-8%	-1%	-9%	-8%	-39%	+9%	-39%	-13%
NewRank	**0.116**	0.004	**0.297**	**0.320**	**0.267**	**0.201**	**5057**	**3113**	**3598**	6638	2050	**4091**
Fullnet NewRank	0.089	**0.020**	0.180	0.191	0.170	0.130	11277	6951	6877	**6541**	**1651**	6659
Gain	-29%	+381%	-66%	-68%	-57%	-54%	-123%	-123%	-91%	+1%	+24%	-63%
YetRank	0.128	0.028	0.206	**0.158**	**0.272**	0.158	2084	**673**	**1048**	846	**492**	**1029**
Fullnet YetRank	**0.157**	**0.029**	**0.224**	0.149	0.259	**0.163**	**2031**	874	1200	**836**	561	1101
Gain	+22%	+1%	+9%	-6%	-5%	+3%	+3%	-30%	-15%	+1%	-14%	-7%

Table 8 Gain of using outsiders as part of the network in the score diffusion step. The *fullnet* versions incorporate outsiders in the network, i.e., they convert outsiders in insiders. Note that OTARIOS does not use outsiders as part of the network in the score diffusion step, only in the initialisation step. The mean of both NDCG and MRR is highlighted, showing that, overall, STOA methods' performance degrades when they use outsiders as insiders.

surprise due to the fact that it presented the worst results in the insiders network. This test demonstrated that adding incomplete information to the citation network (authors whose received citations are unknown), decreases the author rankings if this information is considered on the score diffusion process.

4.2 *Which criteria are more important?*

In order to understand which criteria are more important to create more accurate OTARIOS variants, we analysed the top variants for the NDCG metric. Table 9 shows the criteria considered in the top 20 OTARIOS variants. The results demonstrate that the best variants use a combination of productivity, outsiders influence and score term. Only

5 Conclusions

Here we put forward OTARIOS, a new graph-ranking algorithm to measure authors' scientific impact. Previous graph-ranking algorithms did not combine relevant information effectively, such as the author's productivity and the citations' relevance.

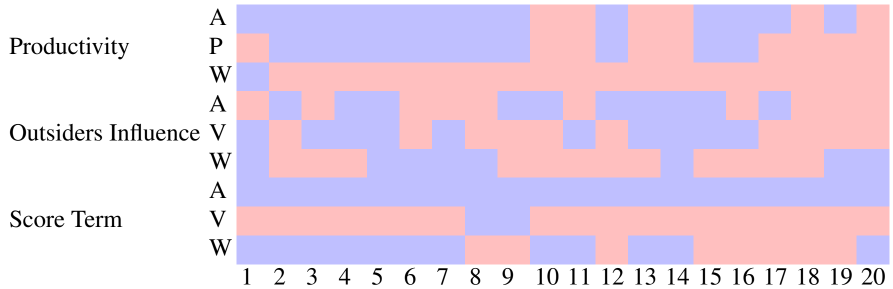

Table 9 Criteria considered on the top 20 OTARIOS variants on the NDCG metric. The rows represent different criteria and the columns the variants that ranked at position *n*. The blue colour in a column indicates that the criteria is considered on the variant, while the red colour indicates its absence.

Furthermore, previous methods assume that the full network is known, which is not true for most real cases. We thus divided the network into insiders (i.e., the authors that we want to rank) and outsiders (i.e., the authors that cite insiders but we do not rank). In our experiments, we analysed which publication/citation information is more relevant and how it can be efficiently combined.

We obtained the best results when OTARIOS considers (i) the author's publication volume and publication recency, (ii) how recently his work is being cited by outsiders, and (iii) how recently his work is being cited by insiders and how individual his work is (i.e., publishing with few authors is better). This evaluation was performed on a set of five networks where the ground-truth was the number of best awards in the conferences belonging to the specific network. Our tests showed that OTARIOS is $\approx 20\%$ more efficient than the best state of the art method (SCEAS) and $\approx 30\%$ more efficient than traditional PageRank (RLPR). We demonstrated that OTARIOS efficiently uses outsiders (i.e., authors whose received citations are not fully known) on the score initialisation process. Furthermore, we showed that adding outsiders to the score diffusion process decreases the performance of state of the art algorithms. These results indicate that current methods have poor results on networks where some nodes have missing information (which is true for most real cases), while OTARIOS is able to use nodes with limited information adequately.

Finally, regarding future work, we plan to test OTARIOS on paper-level citations and verify that we are also capable of improving that approach from the state of the art. Furthermore, we plan to develop a method to automatically identify outsiders (e.g., insiders with low density in the citation network, or insiders with low co-authorship ratio to other insiders) and analyse if this strategy improves author-ranking.

Acknowledgements Jorge Silva is supported by a FCT/MAP-i PhD grant (PD/BD/128157/2016). David Aparício is also supported by a FCT/MAP-i PhD grant (PD/BD/105801/2014).

References

1. B.V., E. *Research Metrics Guidebook*. Elvesier, 2018.
2. DING, Y. Applying weighted pagerank to author citation networks. *Journal of the American Society for Information Science and Technology 62*, 2 (2009), 236–245.
3. DUNAISKI, M., AND VISSER, W. Comparing paper ranking algorithms. In *Proceedings of the South African Institute for Computer Scientists and Information Technologists Conference* (2012), ACM, pp. 21–30.
4. DUNAISKI, M., VISSER, W., AND GELDENHUYS, J. Evaluating paper and author ranking algorithms using impact and contribution awards. *Journal of Informetrics 10*, 2 (2016), 392–407.
5. HIRSCH, J. E. An index to quantify an individual's scientific research output. *Proceedings of the National academy of Sciences 102*, 46 (2005), 16569–16572.
6. HWANG, W.-S., CHAE, S.-M., KIM, S.-W., AND WOO, G. Yet another paper ranking algorithm advocating recent publications. In *Proceedings of the 19th international conference on World wide web* (2010), ACM, pp. 1117–1118.
7. PAGE, L., BRIN, S., MOTWANI, R., WINOGRAD, T., ET AL. The pagerank citation ranking: Bringing order to the web.
8. RADICCHI, F., FORTUNATO, S., MARKINES, B., AND VESPIGNANI, A. Diffusion of scientific credits and the ranking of scientists. *Physical Review E 80*, 5 (2009), 056103.
9. SIDIROPOULOS, A., AND MANOLOPOULOS, Y. Generalized comparison of graph-based ranking algorithms for publications and authors. *Journal of Systems and Software 79*, 12 (2006), 1679–1700.
10. TANG, J., ZHANG, J., YAO, L., LI, J., ZHANG, L., AND SU, Z. Arnetminer: Extraction and mining of academic social networks. In *KDD'08* (2008), pp. 990–998.
11. TANG, J., ZHANG, J., YAO, L., LI, J., ZHANG, L., AND SU, Z. Citation Network Dataset. `https://aminer.org/citation`, 2017. [Online; accessed 14-September-2018].
12. VIEIRA, E. S., CABRAL, J. A., AND GOMES, J. A. How good is a model based on bibliometric indicators in predicting the final decisions made by peers? *Journal of Informetrics 8*, 2 (2014), 390–405.
13. WANG, H., SHEN, H.-W., AND CHENG, X.-Q. Scientific credit diffusion: Researcher level or paper level? *Scientometrics 109*, 2 (2016), 827–837.
14. WEST, J. D., JENSEN, M. C., DANDREA, R. J., GORDON, G. J., AND BERGSTROM, C. T. Author-level eigenfactor metrics: Evaluating the influence of authors, institutions, and countries within the social science research network community. *Journal of the American Society for Information Science and Technology 64*, 4 (2013), 787–801.