Ranking Papers by their Short-Term Scientific Impact

Ilias Kanellos
Athena R.C.
Athens, Greece
ilias.kanellos@athenarc.gr

Thanasis Vergoulis
Athena R.C.
Athens, Greece
vergoulis@athenarc.gr

Dimitris Sacharidis
ULB
Brussels, Belgium
dimitris.sacharidis@ulb.be

Theodore Dalamagas
Athena R.C.
Athens, Greece
dalamag@athenarc.gr

Yannis Vassiliou
NTUA
Athens, Greece
yv@cs.ntua.gr

***Abstract*—The constantly increasing rate at which scientific papers are published makes it difficult for researchers to identify papers that currently impact the research field of their interest. In this work, we present a method that ranks papers based on their estimated short-term impact, as measured by the number of citations received in the near future. Our method models a researcher exploring the paper citation network, and introduces an attention-based mechanism, akin to a time-restricted version of preferential attachment, that explicitly captures the researcher's preference to read papers which received a lot of attention recently. A detailed experimental evaluation on real citation datasets across disciplines, shows that our approach is more effective than previous work.**

***Index Terms*—citation networks, paper ranking, data mining**

I. Introduction

Assessing the scientific impact of publications, colloquially called *papers*, is an important research problem with various applications (e.g., literature exploration, researcher assessment, research funding planing), especially since the number of papers published grows at an increasing rate [1]. Conventionally, quantifying the scientific impact of a paper depends on the network of citations. In this work, we focus on the *short-term impact* (STI) of a paper, quantified by the number of citations it acquires in the near future (referred to as "new citations" [2], or "future citation counts" [3]). Specifically, we address the research problem of ranking papers via their expected STI.

Existing work typically assigns to each paper a proxy score estimating its expected short-term impact. These scores are determined by a stochastic process, akin to PageRank [4], modelling the impact flow in the citation network. The important concern here is to account for the *age bias* inherent in citation networks [5]: as papers can only cite past work, recent publications are at a disadvantage having less opportunity to accumulate citations. A popular way to address this is by introducing time-awareness into the stochastic process, by favoring either recent papers or recent citations [2], [3], [6]. Nonetheless, it has been shown [7] that these methods still leave enough space for further improvements.

In this work we posit that recent citations (i.e., the level of *attention* papers currently enjoy) strongly influence the short-term impact. We investigate this hypothesis and find that it holds to a certain degree across different citation networks. Hence, we introduce an attention-based mechanism, reminiscent of a time-restricted version of preferential attachment [8], that models the fact that recently cited papers continue getting cited in the short-term. We then develop a new paper ranking method, called *AttRank*, that takes advantage of this mechanism to provide improved estimations for ranking papers accoriding to their expected STI.

To evaluate AttRank's effectiveness we perform an extensive experimental evaluation on two citation networks from various scientific disciplines. We measure effectiveness as the ranking accuracy with respect to the ground truth STI ranking. We investigate the importance of the attention mechanism in achieving high effectiveness. We also compare AttRank against several state-of-the-art methods, which are carefully tuned for each experimental setting. Our findings indicate that across almost all settings, AttRank outperforms prior work.

II. Background and Related Work

Citation Network. We represent a collection of papers as a *citation network*, i.e., a directed graph where each node corresponds to a paper, and each edge to a citation. A citation network can be represented by its *citation matrix* $\boldsymbol{C}$, where $C_{i,j} = 1$ iff paper j cites paper i, and $C_{i,j} = 0$, otherwise. We follow the convention that paper ids indicate the publication date order; i.e., paper i was published before j iff $i < j$.

To distinguish between different states of the citation network as it evolves over time, we use the *superscript parenthesis* notation (t) to refer to a state where only papers with id up to t have been published. For example, $\boldsymbol{C}^{(t)}$ is the $t \times t$ citation matrix for the first t papers. Because of how the citation matrix evolves, $\boldsymbol{C}^{(t)}$ is a submatrix of $\boldsymbol{C}^{(t')}$ when $t < t'$. We use the *subscript bracket* notation $[n]$ to refer to a submatrix containing the first n rows and columns, or to a subvector containing the first n entries; $[-n]$ denotes the last n rows/columns/entries. For example, the previous observation can be expressed as $\boldsymbol{C}^{(t)}_{[n]} = \boldsymbol{C}^{(t')}_{[n]}$ for any $n \leq t < t'$.

PageRank. PageRank [4] measures the importance of a node in a network, based on a random walk with jumps process. In the context of citation networks, the process simulates a "random researcher", who starts their work by reading a paper. Then, with probability α, they pick another paper to read from the reference list, or, with probability $1-\alpha$, choose any other paper in the network at random.

Given a citation matrix $\boldsymbol{C}$, we define the stochastic matrix $\boldsymbol{S}$ as follows. Let k_i denote the number of papers cited by i. Then, $S_{i,j} = \frac{1}{k_j}$, iff paper j cites paper i, $S_{i,j} = 0$, iff j does not cite i but cites at least one other paper, and $S_{i,j} = \frac{1}{n}$, iff paper p_j cites no paper (i.e., is a dangling node), where n is the number of papers. Let $\boldsymbol{u}$ denote the *teleport vector* (all vectors are column vectors) such that $|\boldsymbol{u}| \equiv \sum_i u_i = 1$ and $\forall i \; u_i \geq 0$; typically, $\boldsymbol{u}$ is defined to indicate uniform teleport probabilities, i.e., $\forall i \; u_i = 1/n$. Let $\alpha \in [0, 1)$ denote the *random jump probability*. Then the *PageRank vector* $\boldsymbol{v}$ is defined by this equation:

$$\boldsymbol{v} = \alpha \boldsymbol{S} \boldsymbol{v} + (1-\alpha)\boldsymbol{u}. \tag{1}$$

We say that $\boldsymbol{v}$ is the PageRank vector of matrix $\boldsymbol{S}$ with respect to teleport vector $\boldsymbol{u}$. The PageRank vector is given by the following closed-form formula:

$$\boldsymbol{v} = (1-\alpha)\boldsymbol{u} + (1-\alpha)\sum_{x=1}^{\infty} \alpha^x \boldsymbol{S}^x \boldsymbol{u} = (1-\alpha)\sum_{x=0}^{\infty} \alpha^x \boldsymbol{S}^x \boldsymbol{u}, \tag{2}$$

where the convention $\boldsymbol{A}^0 \equiv \boldsymbol{I}$ is used in the last equality. If we define $\boldsymbol{M} = (1-\alpha)\sum_{x=0}^{\infty} \alpha^x \boldsymbol{S}^x$, then we observe the linear relationship between the PageRank and the teleport vector: $\boldsymbol{v} = \boldsymbol{M}\boldsymbol{u}$.

Computing the PageRank vector is not done by computing matrix $\boldsymbol{M}$, but by iteratively estimating $\boldsymbol{v}$ via Equation 1. Starting from some random values for $\boldsymbol{v}$, satisfying $|\boldsymbol{v}| = 1$ and $\forall i \; v_i \geq 0$, at each step we update the PageRank vector as $\boldsymbol{v} \leftarrow \alpha \boldsymbol{S} \boldsymbol{v} + (1-\alpha)\boldsymbol{u}$. In other words, for a given $\boldsymbol{u}$, we have a convenient method to estimate $\boldsymbol{M}\boldsymbol{u}$.

Short-Term Impact (STI). Using node centrality metrics (e.g., PageRank or the number of citations) to capture a paper's impact can introduce biases, e.g., against recent papers, and may render important papers harder to distinguish [9]. This is due to inherent characteristics of citation networks: the references of a paper are fixed, and there is a delay between a paper's publication and its first citation (*"citation lag"* [10]). In contrast, a paper's *short-term impact* [7], also called the number/count of new/future citations [2], [3], looks into a future state of the network and reflects the level of attention (in terms of citations) a paper will receive in the near future.

Consider the state of the citation matrix at present time n. Given a time *horizon* of τ, the *short-term impact* (STI) f_i of a paper i ($i \leq n$) is defined as the number of future citations, i.e., those it would receive in the time period $(n, n+\tau]$:

$$f_i = \sum_{j=n+1}^{n+\tau} C_{i,j}^{(n+\tau)}.$$

Some observations are in order. First, τ is a user-defined parameter that specifies how long in the future one should wait for citations to accumulate. An appropriate value may depend on the typical duration of the research cycle specific to each scientific discipline. Second, it is important to emphasize that STI can only be computed in retrospect; at current time, the future citations are not yet observed. Thus, any method that seeks to identify papers with high STI has to employ a mechanism to account for the unobserved future citations.

With these remarks in mind and similar to prior work [2], [3], [7], we study the following problem.

Problem 1. *Given state $\boldsymbol{C}^{(n)}$ of the citation network at current time n, return a ranking of papers such that it matches their ranking by short-term impact $\boldsymbol{f}$ for a given time horizon τ.*

Impact estimation methods. In recent years, various methods have been proposed for quantifying the scientific impact of papers [7]. A large number of methods are *PageRank adaptations* tailored to better simulate the random researcher behavior (e.g., [11]). However, such approaches do not address age bias. This motivated a number of *time-aware methods*, which introduce time-based weights in the various centrality metric calculations, to favor either recent publications (e.g., [2], [6]) or recent citations (e.g., [3]), or citations received shortly after the publication of an article (e.g., [12]).

Although time-awareness is shown to improve the accuracy when ranking by STI, it fails to differentiate among recent papers favoring all equally. An alternative approach is to combine the basic citation network analysis process with an iterative processes on *multiple interconnected networks*, e.g., author-paper and/or venue-paper networks (indicative example: [13]). There are also methods that consider *ensembles* that combine the rankings from multiple methods. Methods in this category (e.g., [14]) combine several types of scores calculated on various graphs (e.g., citation network, co-authorship network). Finally, a separate line of work is concerned with *impact prediction algorithms* which are based on modeling the arrivals of citations for individual papers (e.g., [15]), however such methods are prone to overfitting [16], and require a long history (≥ 5 years) of observed citations for each paper.

It is apparent that, all previous approaches focus on using additional data resources (e.g., venues, co-authorship networks) in an attempt to alleviate the aforementioned recent paper differentiation issue of the time-aware methods. However, such data is not readily available, fragmented in different datasets, not easy to collect, integrate and clean, and is often incomplete. In contrast, our approach is to rely solely on the properties of the underlying citation network, and try to better model the process with which the network evolves.

III. Approach

Overview. To rank by short term impact without knowing the future, one needs to introduce assumptions about how the citation network evolves. In this work, we assume that the PageRank vector at time $t \geq n$ indicates the chance the papers will get cited at future time $t+1$. This defines a mechanism, where a random researcher explores literature. While reading a paper, the researcher may choose with probability α to further read one of the paper's references, or pick another paper at random from some prior (teleport) probability. So the probability of a paper i receiving a citation from a random researcher at time $t+1$ is proportional to their PageRank $v_i^{(t)}$ at time t, which satisfies:

$$\boldsymbol{v}^{(t)} = \alpha \boldsymbol{S}^{(t)} \boldsymbol{v}^{(t)} + (1-\alpha)\boldsymbol{u}^{(t)},$$

where $\boldsymbol{S}^{(t)}$ is the stochastic matrix and $\boldsymbol{u}^{(t)}$ is the teleport vector at t.

The short-term impact of a paper is the sum of citations it will receive at each time t between $n+1$ and $n+\tau$. Let us further assume that each paper in the future (after current time n) makes the same number of citations. So, under our assumptions, the number of future citations f_i a paper i receives is proportional to:

$$f_i \propto \sum_{t=0}^{\tau-1} v_i^{(n+t)}.$$

For the purposes of ranking, the scale of individual f_i does not matter. Therefore, to rank by STI we would like to estimate the following vector:

$$\boldsymbol{y} = \sum_{t=0}^{\tau-1} \boldsymbol{v}_{[n]}^{(n+t)}, \tag{3}$$

i.e., the sum at each future timestamp of the PageRank of the present n papers.

In Equation 3, each PageRank vector $\boldsymbol{v}_{[n]}^{(n+t)}$ is a random vector under the aforementioned citation network growth process, and its value depends on the values of the PageRank vector at previous times. Therefore, one way to estimate the mean of $\boldsymbol{y}$ is with Markov Chain Monte Carlo methods to draw samples from the probability distribution of $\boldsymbol{y}$. This however is costly, as each sample requires the computation of τ PageRank vectors for a large citation network with at least n nodes.

We propose a different approach. We start by assuming that the future citation matrix and thus $\boldsymbol{S}^{(n+\tau)}$ is known. Conditional to $\boldsymbol{S}^{(n+\tau)}$, the PageRank vector $\boldsymbol{v}_{[n]}^{(n+t)}$ become independent. We then apply a mechanism to rewrite Equation 3 so that it can be computed with a single PageRank computation. This rewriting now has quantities derived from the $\boldsymbol{S}^{(n+\tau)}$ matrix. At the final step, we estimate these quantities from the current (at time n) state of the network, and compute the rewritten Equation 3.

We next describe the mechanism that allows us to rewrite Equation 3 as a single PageRank computation.

Contraction. Consider a network of $n+m$ nodes, represented by its stochastic matrix $\boldsymbol{S}$, where its m last nodes have no incoming edges. Given some teleport vector $\boldsymbol{u}$ and random jump probability α, let $\boldsymbol{v}$ denote the PageRank vector that satisfies $\boldsymbol{v} = (1-\alpha)\sum_{x=0}^{\infty} \alpha^x \boldsymbol{S}^x \boldsymbol{u}$. The PageRank scores $\boldsymbol{v}_{[n]}$ of the first n nodes can be computed directly from the PageRank on the stochastic matrix $\boldsymbol{S}_{[n]}$ with respect to an adjusted teleport vector $\mathring{\boldsymbol{u}}$, as indicated by the following theorem. Note that all proofs can be found in the extended version of this paper [17].

Theorem 1. *Let $\boldsymbol{v}$ denote the PageRank vector for stochastic matrix $\boldsymbol{S} = \begin{pmatrix} \boldsymbol{S}_{[n]} & \boldsymbol{S}_{[n],[-m]} \\ \boldsymbol{0}_{(m\times n)} & \boldsymbol{0}_{(m\times m)} \end{pmatrix}$ with respect to teleport vector $\boldsymbol{u}$ and random jump probability α. Define the adjusted teleport vector $\mathring{\boldsymbol{u}}_{[n]} = \boldsymbol{u}_{[n]} + \alpha \sum_{i=1}^{m} u_{n+i} S_{[n],n+i}$. Then it holds that:*

$$\frac{1}{|\boldsymbol{v}_{[n]}|}\boldsymbol{v}_{[n]} = \frac{\alpha}{|\boldsymbol{v}_{[n]}|}\boldsymbol{S}_{[n]}\boldsymbol{v}_{[n]} + \frac{1-\alpha}{|\mathring{\boldsymbol{u}}_{[n]}|}\mathring{\boldsymbol{u}}_{[n]},$$

where $|\mathring{\boldsymbol{u}}_{[n]}| = |\boldsymbol{v}_{[n]}| = 1-(1-\alpha)\sum_{i=1}^{m} u_{n+i}$.

In other words, to compute the PageRank w.r.t. $\boldsymbol{S}$ for the first n nodes, we can use an adjusted teleport vector $\mathring{\boldsymbol{u}}_{[n]}$ (after normalization) to compute the PageRank w.r.t. $\boldsymbol{S}_{[n]}$, and then scale the result by $|\mathring{\boldsymbol{u}}_{[n]}|$.

Applying Contraction. We will apply the contraction idea to compute each future PageRank vector $\boldsymbol{v}_{[n]}^{(n+t)}$ as a PageRank vector of the current stochastic matrix $\boldsymbol{S}^{(n)}$, assuming that its future state $\boldsymbol{S}^{(n+t)}$ is known. Note that to apply the contraction idea, we need to restrict each future paper i to cite no other future paper j ($n < j < i$), i.e., citations only come for papers already published until current time n. This restriction only affects the PageRank values of the future papers, which however we do not need to rank.

Note that for any time $n+t$ where $t > 0$, the first n rows and columns of the stochastic matrix $\boldsymbol{S}^{(n+t)}$ remain constant, and we denote $\boldsymbol{S} \equiv \boldsymbol{S}_{[n]}^{(n+t)}$. From Theorem 1 we have:

$$\frac{1}{|\boldsymbol{v}_{[n]}^{(n+t)}|}\boldsymbol{v}_{[n]}^{(n+t)} = \frac{\alpha}{|\boldsymbol{v}_{[n]}^{(n+t)}|}\boldsymbol{S}\boldsymbol{v} + \frac{1-\alpha}{|\boldsymbol{v}_{[n]}^{(n+t)}|}\mathring{\boldsymbol{u}}_{[n]}^{(n+t)}.$$

Defining $\boldsymbol{M} = (1-\alpha)\sum_{x=0}^{\infty}\alpha^x \boldsymbol{S}^x$, we rewrite the previous equation in the closed form of Equation 2:

$$\boldsymbol{v}_{[n]}^{(n+t)} = \boldsymbol{M}\mathring{\boldsymbol{u}}_{[n]}^{(n+t)}.$$

From the definition of STI, we derive:

$$\boldsymbol{y} = \sum_{t=0}^{\tau-1} \boldsymbol{v}_{[n]}^{(n+t)} = \boldsymbol{M}\sum_{t=0}^{\tau-1}\mathring{\boldsymbol{u}}_{[n]}^{(n+t)},$$

where $\mathring{\boldsymbol{u}}_{[n]}^{(n+t)} = \boldsymbol{u}_{[n]}^{(n+t)} + \alpha\sum_{i=1}^{m} u_{n+i}^{(n+t)} S_{[n],n+i}^{(n+t)}$.

For convenience, we further assume that the teleport vector for the first n papers is the same at each time $n+t$, and we denote it as $\boldsymbol{u} \equiv \boldsymbol{u}_{[n]}^{(n+t)}$. We thus split the adjusted teleport vectors into a non-time-dependent component and a time-dependent component: $\mathring{\boldsymbol{u}}_{[n]}^{(n+t)} = \boldsymbol{u} + \boldsymbol{w}_{[n]}^{(n+t)}$. Summing the time-dependent components for $0 \le t \le \tau-1$, we introduce:

$$\boldsymbol{w} \equiv \sum_{t=0}^{\tau-1}\boldsymbol{w}_{[n]}^{(n+t)} = \alpha\sum_{t=0}^{\tau-1}\sum_{i=1}^{t} u_{n+i} S_{[n],n+i}^{(n+t)}.$$

Then, the STI can be expressed as:

$$\boldsymbol{y} = \boldsymbol{M}(\tau\boldsymbol{u} + \boldsymbol{w}),$$

meaning that $\hat{\boldsymbol{y}} \equiv \frac{\boldsymbol{y}}{|\boldsymbol{y}|}$ can be computed as the PageRank vector of matrix $\boldsymbol{S}$ with respect to teleport vector $\frac{\tau}{|\boldsymbol{y}|}\boldsymbol{u} + \frac{1}{|\boldsymbol{y}|}\boldsymbol{w}$. Introducing coefficients $\alpha, \beta, \gamma \in [0,1]$, such that $\alpha+\beta+\gamma = 1$, we can write STI in a general form:

$$\hat{\boldsymbol{y}} = \alpha\boldsymbol{S}\hat{\boldsymbol{y}} + \beta\hat{\boldsymbol{w}} + \gamma\hat{\boldsymbol{u}}, \tag{4}$$

where $\hat{\boldsymbol{w}}$ and $\hat{\boldsymbol{u}}$ are the normalized vectors of $\boldsymbol{w}$ and $\boldsymbol{u}$.

Attention. Because the time-dependent vector $\boldsymbol{w}$ is determined by quantities of future states $\boldsymbol{S}^{(n+t)}$, we need to estimate it from the known current state $\boldsymbol{S}$. A simple way is instead of going τ time steps in the future, to go y time steps in the past. Assuming teleport probabilities u_{n+t} for future papers are equal, we estimate:

$$\tilde{\boldsymbol{w}} \propto \sum_{t=0}^{y-1} \sum_{i=0}^{t} S_{[n],n-i} = \sum_{i=0}^{y-1} (y-i) S_{[n],n-i}. \tag{5}$$

We call this estimated vector $\tilde{\boldsymbol{w}}$ the *attention* vector, because for each paper i, it computes a weighted count of its citations from the y most recent papers, i.e., its recent attention $\tilde{w}_i \propto yS_{i,n} + (y-1)S_{i,n-1} + \cdots + S_{i,n-y+1}$.

Attention, however, is not the only mechanism that governs which papers researchers read. Naturally, researchers may read a paper cited in the reference list of another paper. Moreover, similar to previous work [2], [6], we assume that researchers also read recently published papers. Specifically, we capture the recency of a paper i using an exponentially decaying score:

$$u_i = c e^{\eta \cdot (n-i)}, \tag{6}$$

where n is the current time, $i < n$ denotes the publication time of paper i, hyperparameter η is a negative constant (as $n - i \geq 0$), and c is normalization constant so that $|\boldsymbol{u}| = 1$. To calculate a proper η value, a similar procedure like the one used in [6] can be followed (see also Section IV-B).

AttRank. We refer to the ranking approach based on Equations 4, 5, and 6 as AttRank. Note that Eq. 4 combines three mechanisms. Specifically, we assume that researchers read a paper for one of the following reasons: the paper gathered attention recently, was recently published, or was found in another paper's reference list. We model this behavior with the following random process. A researcher chooses to read any other paper from paper i's reference list, after reading it, with probability α. With probability β she chooses a paper based on its attention. This behavior makes recently rich papers even richer, and is reminiscent of a time-restricted preferential attachment mechanism of the Barabási-Albert network growth model [8]. Finally, with probability γ she chooses any paper with a preference towards recently published ones.

Two special values for coefficient β are noteworthy. First, observe that when $\beta = 0$, a setting we call NO-ATT (for no attention), the model becomes similar to time-aware methods that address the inherent bias against new papers in citation networks (see [7] for a thorough coverage of such approaches). Note that additionally setting $\eta = 0$ in Eq. 6 recovers PageRank. Second, when $\beta = 1$, a setting we call ATT-ONLY (for attention only), AttRank is solely based on the attention mechanism, assuming that the recent citation patterns will persist exactly in the near future. To the best of our knowledge, ATT-ONLY has not been considered in the literature as a means to estimate the short-term impact of a paper. As we show in Section IV, attention alone is a powerful mechanism, often outperforming existing approaches. However, $\beta = 1$ is never the optimal setting; it is always better to consider attention in combination with the other two citation mechanisms.

Equation 4 describes an iterative process for estimating STI vector: starting with a random value, at each step update the vector with the right hand side of the equation. This process is repeated until the values converge. The following theorem, ensures that convergence is achievable.

Theorem 2. *The iterative process defined by Eq. 4 converges.*

IV. Evaluation

This section presents an experimental evaluation of our approach for ranking papers based on their short-term impact. Specifically, Section IV-A discusses the experimental setup and evaluation approach taken. Section IV-B investigates the effectiveness of our proposed method and the importance of the attention-based mechanism. Section IV-C compares AttRank with existing approaches from the literature.

A. Experimental Setup

Datasets. We consider two datasets in our experiments:

1) American Physical Society (APS)[1] dataset, which contains about $500,000$ papers, written from 1893 to 2014.
2) *DBLP*,[2] which contains about 3 million computer science papers published from 1936 to 2018.

Evaluation Methodology. To evaluate the effectiveness of AttRank in ranking papers based on their short-term impact, we construct a *current* and a *future* state of the citation network. We partition each dataset according to time in two parts, each having an equal number of papers. We use the older half to construct the current state of the citation network, denoted as $\boldsymbol{C}^{(n)}$. All ranking methods will be based on this network acting as the "training" subset. We use parts of the newer half to construct the future state of the network, denoted as $\boldsymbol{C}^{(n+\tau)}$. All ranking methods will be evaluated based on this network acting as the "test" subset.

Specifically, the future state is constructed as follows: we vary the size, in terms of number of papers, of the future state relative to the size of the current state. Thus we do not vary the time horizon τ directly, but rather the *test ratio*, which is the relative size of the future with respect to the current network. We consider values for the test ratio among $\{1.2, 1.4, 1.6, 1.8, 2.0\}$, where 2.0 corresponds to using all citations in the dataset to define the future state. In some experiments we fix the test ratio to a default value of 1.6, meaning that the future state contains 30% more papers than the current state. Note, that the relationship between test ratio and τ is not linear, due to the non-constant number of published papers per year and the fact that most datasets contain incomplete entries for the last year they include.

Given the future state of the citation network, we can compute the STI of each paper as per its definition (see Section II). Similar to previous approaches [2], [3], [6], [7],

[1]https://journals.aps.org/about
[2]https://aminer.org/citation

the ranking of papers based on their STI forms the *ground truth*. Any paper ranking method is oblivious of the future state $\boldsymbol{C}^{(n+\tau)}$ of the citation network, and hence the ground truth, and only uses the current state $\boldsymbol{C}^{(n)}$ to derive a ranking. To quantify the effectiveness of a method, we compare its produced ranking to the ground truth, using Spearman's ρ [18], and the Normalized Discounted Cumulative Gain at rank k (nDCG@k). While Spearman's ρ calculates an overall similarity of the given ranking with the ground truth, nDCG@k measures the agreement of the two rankings on the top-ranking papers. In our evaluation, we consider values of k among $\{5, 10, 50, 100, 500\}$, with $k = 50$ being the default value.

B. Ranking Effectiveness

In this section, we investigate AttRank's effectiveness for the default experimental setting (test ratio equal to 1.6), varying parameters, α, β, γ, and y. We vary α, β, γ with a step of 0.1 in the ranges $[0.0, 0.5]$,$[0.0, 1.0]$, $[0.1, 0.9]$, respectively and y in $\{1, 2, 3, 4, 5\}$. For each metric, we discuss AttRank's parameterization that achieves the best ranking effectiveness.

First, however, we discuss how we set the value of the exponential factor η of Equation 6. We follow a similar approach as the one used in [6]. For each dataset, we use an exponential function of the form $e^{\widetilde{\eta}y}$, to fit the tail of the empirical distribution of the random variable Y that models the probability of an article being cited n years after its publication. The factor $\widetilde{\eta}$ of the fitting function is used as the η value. Following this procedure, we calculate $\eta = -0.12$ for APS and $\eta = -0.16$ for DBLP.

1) Effectiveness in terms of Correlation: In this experiment, we measure AttRank's ranking effectiveness in terms of Spearman's ρ to the ground truth ranking by STI. We find the optimal values for $\{\alpha, \beta, \gamma, y\}$ at $\{0.3, 0.3, 0.4, 3\}$ for APS ($\rho = 0.6295$) and at $\{0.2, 0.4, 0.4, 3\}$ for DBLP ($\rho = 0.6316$). Overall, we find that AttRank correlates at least moderately to the STI ranking in its best setting. Further, the significance of the attention mechanism is evident since the best correlation is achieved when $\beta > 0$. Finally, the maximum correlation values for $\beta = \{0, 1\}$ are $\{0.581, 0.537\}$, for APS, and $\{0.529, 0.571\}$ for DBLP, respectively.

2) Effectiveness in terms of nDCG@50: We repeat the effectiveness analysis, this time considering the nDCG@50 metric. We determine the parameterization that achieves the best nDCG@50 per dataset and we find that it corresponds to $\{\alpha$,β,γ,$y\}$ at $\{0.3$,0.5,0.2,$3\}$ for APS ($nDCG = 0.7293$), and $\{0.0,\ 0.1,\ 0.9,\ 1\}$ for DBLP ($nDCG = 0.9168$). Again, we observe that the attention vector plays a non-negligible role in achieving the maximum nDCG ($\beta > 0$). Indicatively, the maximum nDCG@50 values for $\beta = \{0, 1\}$ are $\{0.635, 0.709\}$ for APS, and $\{0.663, 0.916\}$ for DBLP, respectively.

C. Comparative Evaluation

In this section, we compare AttRank to other approaches for impact-based paper ranking. We select the five methods found to be most effective in ranking by short-term impact in [7]:

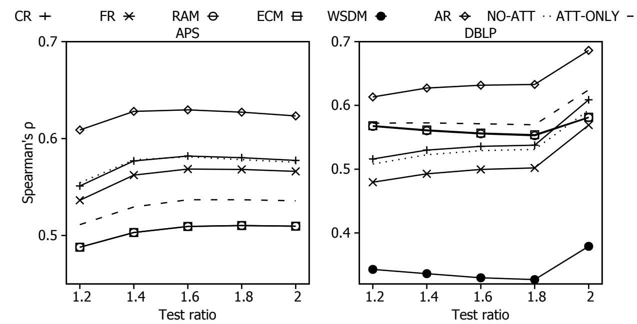

Fig. 1: Effectiveness of all methods in terms of correlation.

- **CiteRank (CR)**. This PageRank-based method [2] calculates the "traffic" towards papers by researchers that prefer reading recent papers.
- **FutureRank (FR)**. This time-aware method [6] combines PageRank and HITS applying mutual reinforcement from papers to authors and vice versa.
- **Retained Adjacency Matrix (RAM).** This citation count variant uses a citation age-weighted adjacency matrix [3].
- **Effective Contagion Matrix (ECM).** This method operates over a citation age-weighted adjacency matrix [3] and calculates weights of citation chains.
- **WSDM cup's 2016 winner (WSDM)**. This method [14] calculates paper scores as combinations of scores propagated by their authors, venues, and citing papers.

The optimal parameterization for the competitors, presented in each individual work, results from the use of particular datasets and experimental settings, which differ among them. Therefore, we extensively tuned all competitors, to ensure a fair comparison of their effectiveness in ranking based on STI.[3] We ran all iterative methods with a convergence error of 10^{-12}, to ensure that further iterations are not expected to change the rankings. We also present the NO-ATT, and ATT-ONLY variants ($\beta = 0$, and $\beta = 1$), to demonstrate the effect of the attention mechanism. Note, that we did not run WSDM on APS, since it lacks the venue data required by the method.

Figure 1 presents the correlation of each method's ranking to that of the STI ranking. We vary the test ratio of the size of networks according to Section IV-A. For each setting we choose the parameterization with the best correlation. We observe that AttRank better correlates to the STI ranking, compared to all competitors increasing correlation by up to 0.057 and 0.079 compared to the best competitor, on APS and DBLP, respectively. AttRank's performance is due to the fact that it does not simply promote papers recently cited, or published, as its competitors do. Instead, its attention mechanism promotes well-cited, recent papers, compared to lesser cited recent papers. Moreover, AttRank promotes older papers that are still heavily cited. The attention mechanism's importance is illustrated by ATT-ONLY's performance, which is the second best on DBLP. Most importantly, AttRank's effectiveness is

[3]For a complete description of each method's parameters and examined settings in our evaluation, refer to the extended version of this paper [17].

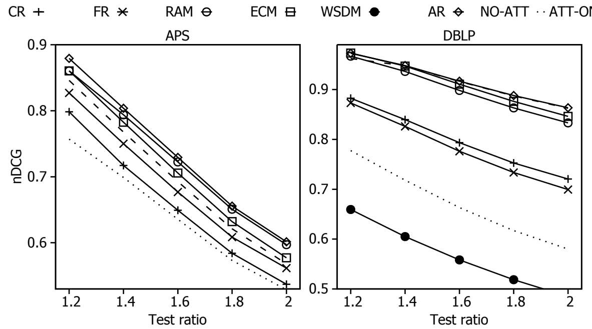

Fig. 2: Effectiveness in terms of nDCG@50.

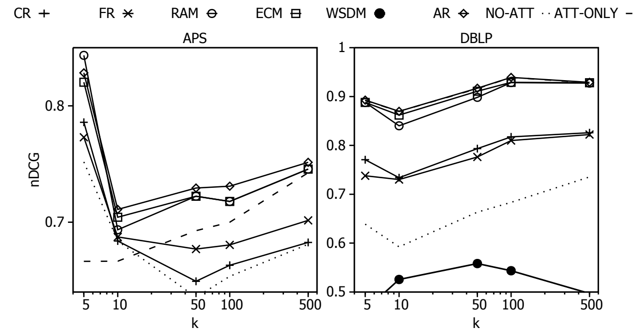

Fig. 3: Effectiveness in terms of nDCG@k (test ratio at 1.6).

always increased when the attention mechanism is balanced with AttRank's other mechanisms.

We measure the effectiveness based on the nDCG achieved by each method with two experiments: in the first, we measure nDCG varying the test ratio for $k = 50$. In the second experiment we use the default test ratio (at 1.6) and measure nDCG varying k. Figures 2 and 3 present the respective results. In general we observe that AttRank is at least on par, and mostly outperforms all rivals, losing only to ECM on DBLP for the test ratio at 1.2 in the first experiment (albeit with a difference less than 0.001) and to RAM for $k = 5$ in the second experiment (diff. less than 0.1). Overall, AttRank improves the nDCG@50 by up to 0.018 and 0.017 on APS and DBLP respectively, while it improves the nDCG@k by up to 0.017 and 0.01 on APS and DBLP, respectively. An interesting observation from the first experiment is that as we look further into the future (i.e., the test ratio) increases, the nDCG achieved by all methods tends to decrease. Overall, the best rival methods in both scenarios are RAM and ECM.

Regarding AttRank's special cases, we observe in both Figures 2 and 3, that excluding attention (NO-ATT) results in a significant drop in nDCG. On the other hand, attention alone (ATT-ONLY) can outperform many existing methods. As also shown in the correlation experiment, AttRank achieves the best results when it combines attention with its other mechanisms.

V. Conclusion

In this work, we present AttRank, a method that effectively ranks papers based on their expected short-term impact. The key idea is to utilize the recent attention a paper has received. Our method models the process of a random researcher reading papers from the literature, and incorporates an attention mechanism to identify popular papers that are likely to continue receiving citations. We studied the effectiveness of our approach in terms of Spearman's rank correlation and nDCG compared to the STI rankings of papers across different citation networks. Our findings validate the introduction of the attention-based mechanism and demonstrate that our method outperforms existing methods in terms of both metrics.

References

[1] L. Bornmann and R. Mutz, "Growth rates of modern science: a bibliometric analysis based on the number of publications and cited references," *Journal of the Association for Information Science and Technology*, vol. 66, no. 11, pp. 2215–2222, 2015.

[2] D. Walker, H. Xie, K.-K. Yan, and S. Maslov, "Ranking scientific publications using a model of network traffic," *Journal of Statistical Mechanics: Theory and Experiment*, vol. 2007, no. 06, p. P06010, 2007.

[3] R. Ghosh, T.-T. Kuo, C.-N. Hsu, S.-D. Lin, and K. Lerman, "Time-aware ranking in dynamic citation networks," in *Data Mining Workshops (ICDMW), 2011 IEEE 11th International Conference on*. IEEE, 2011, pp. 373–380.

[4] L. Page, S. Brin, R. Motwani, and T. Winograd, "The pagerank citation ranking: Bringing order to the web." Stanford InfoLab, Tech. Rep., 1999.

[5] H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, and M.-Y. Zhou, "Ranking in evolving complex networks," *Physics Reports*, vol. 689, pp. 1–54, 2017.

[6] H. Sayyadi and L. Getoor, "Futurerank: Ranking scientific articles by predicting their future pagerank," in *Proceedings of the 2009 SIAM International Conference on Data Mining*. SIAM, 2009, pp. 533–544.

[7] I. Kanellos, T. Vergoulis, D. Sacharidis, T. Dalamagas, and Y. Vassiliou, "Impact-based ranking of scientific publications: a survey and experimental evaluation," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[8] A.-L. Barabási, "Network science book," *Boston, MA: Center for Complex Network, Northeastern University. Available online at: http://barabasi.com/networksciencebook*, 2014.

[9] W.-S. Hwang, S.-M. Chae, S.-W. Kim, and G. Woo, "Yet another paper ranking algorithm advocating recent publications," in *Proceedings of the 19th international conference on World wide web*. ACM, 2010, pp. 1117–1118.

[10] V. P. Diodato and P. Gellatly, *Dictionary of Bibliometrics (Haworth Library and Information Science)*. Routledge, 1994.

[11] P. Chen, H. Xie, S. Maslov, and S. Redner, "Finding scientific gems with google's pagerank algorithm," *Journal of Informetrics*, vol. 1, no. 1, pp. 8–15, 2007.

[12] E. Yan and Y. Ding, "Weighted citation: An indicator of an article's prestige," *Journal of the Association for Information Science and Technology*, vol. 61, no. 8, pp. 1635–1643, 2010.

[13] X. Jiang, X. Sun, and H. Zhuge, "Towards an effective and unbiased ranking of scientific literature through mutual reinforcement," in *Proceedings of the 21st ACM international conference on Information and knowledge management*. ACM, 2012, pp. 714–723.

[14] M. Feng, K. Chan, H. Chen, M. Tsai, M. Yeh, and S. Lin, "An efficient solution to reinforce paper ranking using author/venue/citation information-the winner's solution for wsdm cup 2016," *WSDM Cup*, 2016.

[15] R. Yan, J. Tang, X. Liu, D. Shan, and X. Li, "Citation count prediction: learning to estimate future citations for literature," in *Proceedings of the 20th ACM international conference on Information and knowledge management*. ACM, 2011, pp. 1247–1252.

[16] J. Wang, Y. Mei, and D. Hicks, "Comment on "quantifying long-term scientific impact"," *Science*, vol. 345, no. 6193, pp. 149–149, 2014.

[17] I. Kanellos, T. Vergoulis, D. Sacharidis, T. Dalamagas, and Y. Vassiliou, "Ranking papers by their short-term scientific impact," *arXiv preprint arXiv:2006.00951*, 2020.

[18] C. Spearman, "The proof and measurement of association between two things," *The American journal of psychology*, vol. 15, no. 1, pp. 72–101, 1904.