IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received 2 June 2023, accepted 13 June 2023, date of publication 28 June 2023, date of current version 10 August 2023.

Digital Object Identifier 10.1109/ACCESS.2023.3290906

RESEARCH ARTICLE

Predicting the Future Popularity of Academic Publications Using Deep Learning by Considering It as Temporal Citation Networks

KHUSHNOOD ABBAS[1,2], MOHAMMAD KAMRUL HASAN[3], (Senior Member, IEEE), ALIREZA ABBASI[2], UMI ASMA MOKHTAR[3], ASIF KHAN[4],(Member, IEEE), SITI NORUL HUDA SHEIKH ABDULLAH[3], SHI DONG[1], (Member, IEEE), SHAYLA ISLAM[5], (Senior Member, IEEE), DABIAH ALBOANEEN[6], AND FATIMA RAYAN AWAD AHMED[7]

[1]School of Computer Science and Technology, Zhoukou Normal University, Zhoukou, Henan 466000, China
[2]School of Engineering and IT, University of New South Wales (UNSW), Sydney, NSW 2052, Australia
[3]Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia (UKM), Bangi, Selangor 43600, Malaysia
[4]School of Computer Application, Integral University, Lucknow 226026, India
[5]Institute of Computer Science and Digital Innovation, UCSI University, Kuala Lumpur 56000, Malaysia
[6]Computer Science Department, College of Sciences and Humanities, Imam Abdulrahman Bin Faisal University, Al Jubail 34212, Saudi Arabia
[7]Computer Science Department, College of Computer Engineering and Science, Prince Sattam Bin Abdulaziz University, Al-Kharj 16273, Saudi Arabia

Corresponding authors: Mohammad Kamrul Hasan (hasankamrul@ieee.org) and Asif Khan (asif05amu@gmail.com)

This work was supported in part by the Ministry of Higher Education Malaysia through the Research Grant Scheme under Grant FRGS/1/2020/ICT03/UKM/02/6.

ABSTRACT One of the key goals of Informetrics is to identify citation-based popular articles among so many other aspects, such as determining popular research topics, identifying influential scholars, and predicting hot trends in science. These can be achieved by applying network science approaches to scientific networks and formulating the problem as a popular (most-cited) node ranking task. To rank the papers based on their future citation gain. In this work a deep learning based framework is proposed. Which helps in automatic node level feature extraction and can make node level prediction in dynamic graphs such as citation networks. To achieve this we have learned global ranking preserve d dimensional node embedding. We have only considered temporal features, which makes it suitable for generalisation to other networks. Although our model can consider node level explicit features also. Further we have given novel cost function which can be easily solve ranking problem for dynamic graphs using probabilistic regression method. Which can be easily optimised. Another novelty of our work is that our model can be trained using different snapshots of the graph and different time. Further trained model can be used to make future prediction. The proposed model has been tested on an arXiv paper citation network using six standard information retrieval-based metrics. The results show that our proposed model outperforms, on average, other state-of-the-art static models as well as dynamic node ranking models. The outcome of this research study leads to informed data-driven decision-making in science, such as the allocation and distribution of research funds and investment in strategic research centers. When considering past time window size as 10 months and making prediction after 10 months our proposed model's performance on various ranking based evaluation metrics are as follows: AUC-0.974, Kendal's rank correlation tau-0.455, Precision- 0.643, Novelty-0.0456, Temporal novelty-0.375 and on NDCG-0.949. Our model is able to make long term trend prediction with just training on short time window.

INDEX TERMS Citation prediction, citation networks, node ranking, deep learning, temporal networks, and popularity prediction.

The associate editor coordinating the review of this manuscript and approving it for publication was Yiqi Liu.

I. INTRODUCTION

Scholars and scholarly artifacts such as research publications, research groups, journals, and research topics have been an active sub-field of Scientometrics and Informetrics. Such analysis helps to devise appropriate decisions and/or policies about how research funds can be allocated and the way scientists are hired. Network science has been utilised in this domain by formulating this task as a ranking node problem in temporal scholarly networks such as co-authorship, citation, or co-word networks. This approach has been adopted in a broad range of real-world problems [1], [2], [3] such as the way we find the information and products online [4], [5] identifying popular articles, trend prediction in science [6] and Nobel laureates [7] determining users' choices when buying items, watching videos, reading articles [8], [9], [10] trending topics/news in social media [8], [11] disease propagation control, cyber-security, and financial and political activities on social media.

Accomplishing such comprehension turns out to be increasingly difficult as the amount of user-generated data increases over time, and therefore, more advanced data-centric techniques such as machine learning are needed to effectively process the huge volume of data.

Currently, the Scientometrics and Informetrics community focuses on ranking problems in the newly evolving field of ''Science of Science'' [12], [13]. Citation counts are considered a good proxy for measuring the impact of scientific papers or authors, and despite having drawbacks [14] such as the fact that not all citations carry the same weight [15], [16], citation counts can be manipulated [17]. A citation could exist in background reading, crediting, validating, correcting, or even criticizing (i.e., citing in negative ways) [18], [19], [20]. There exist many factors that cannot be easily considered. Classical citation-based methods often only consider the number of citations received by an article or a simple equation like the average for authors and journals. However, network-based models consider the entire structure of the academic networks (e.g., citation and co-citation based) and the position of the nodes (e.g., publications, authors, journals) in such networks to identify the most important or central nodes.

PageRank [21] algorithm considers the weight of citations: if a paper is cited by a highly cited paper, then it will receive a higher ranking score. But the PageRank algorithm has some weaknesses, such as being biased toward old nodes. In other words, highly cited papers will be ranked higher over time, even if they are of no use.

The ranking problem can be solved using a network or graph representation of the data, particularly for heterogeneous data types. One of the best data structures to store such data heterogeneity just emerged as graphs or networks. Graphs are also the greatest option for illustrating the link between entities, whether it is implicit or explicit. Networks are composed of a set of nodes V and a set of edges E reflecting a connection between each pair of nodes, and therefore, can be represented as a set of two tuples $G(V, E)$. However, the corresponding edges of our models of the actual interactions in our daily lives have a timestamp T. So, a dynamic or temporal graph G can be represented by a three-tuples set $G(V, E, T)$.

This paper aims to develop an efficient method to identify the citation-based popularity of academic publications and predict their popularity over time using a machine learning-based method that incorporates various features of big temporal data. This prediction problem has been translated to a node ranking problem in temporal academic networks. An advanced machine learning-based approach is applied to represent a low-dimensional knowledge graph [22] to learn nodes' useful features. Considering the use of a network-based approach and the wide applicability of networks in our real life, the proposed method can be utilized for ranking items to identify their popularity in other domains, such as the content on social media platforms, online items in e-commerce applications, and modeling interactions among different items. To achieve this, we consider a graph snapshot from a previous time point and extract graph-related features at every point. To make our problem more machine learning-friendly, we further convert the ranking problem into a probabilistic regression problem. We have achieved this by calculating the total citation gains and normalizing them by dividing the total number of citation gains. Further, we sampled it from the cumulative distribution of the citation gains. Our key research question is: 'How to identify and predict the (citation-based) popularity of academic publications over time?'. The primary contribution of this study is to use Ml to address this difficulty in the field of informetrics by creating a cutting-edge deep learning-based algorithm for citation prediction that takes into account articles' rank on their temporal citation networks.

The main contributions to this work are listed below:

1) We took the academic paper's future prediction problem and formulated it as a ranking node in temporal or dynamic networks.
2) To solve the ranking problem, we have developed an automated feature extraction method. which needs node-level temporal details only, which makes our method applicable for large-scale networks. Our model uses nodes' local information and predicts nodes' global scores.
3) Furthermore, we used the neural network to solve our problem as a probabilistic regression problem that suits the learning problem in ranking scenarios.
4) We have performed tests on real paper citation networks and found that while learning, we need a little time, but our model can predict future rankings.

The organization of this paper is as follows. In Section II, we review previous work on ranking and related topics. In Section III, we present our proposed method and describe the benchmark methods. In Section III-C, we describe the citation dataset used in our study. We discuss the experimental setup in Section IV. We present the experimental results and analysis in SectionIV-E. We go over the model's performance

and other issues in Section VI. We conclude up the study by going over its potential real-time applications, drawbacks, and future work in Section VI.

II. LITERATURE REVIEW

There are various classical citation-based methods in the fields of scientometrics and Informetrics for ranking academic publications (i.e., h-index, g-index) and other entities such as journals (e.g., impact factor). However, reviewing those studies is beyond the scope of this work. For more details, we suggest interested readers consult review publications like [23] and [24]. Considering the main focus of this study is to present our proposed ML-based method in Informetrics, in this section we mainly review the relevant literature on citation prediction approaches using node ranking approaches in static as well as dynamic graphs in network science, after a brief review of the key features and predictors for citation analysis.

A. NODE RANKING IN STATIC GRAPHS

Node ranking models or algorithms in static graphs generally aim to find the most important or central nodes in networks. These kinds of models or algorithms take the whole network topology into account but often ignore the temporal aspect of the network. The most basic static centrality measures that are widely used are, node degree (or in-degree), betweenness centrality [25], Katz centrality [26]. The degree, also known as in-degree in directed networks, is the fundamental measure of assessing the importance of nodes in a network. The degree centrality considers a node is important if it has many connections, meaning it only considers the direct links to a node. Although the degree of centrality does not take the overall topology of the network into consideration, it is a relevant measure for some systems. Node degree can be a good measure if the system is driven by the rich-get-richer effect [27] but in reality, evolving networks such as citation networks are driven by many processes [4], [28]. Therefore, researchers are proposing alternative approaches based on other measures of importance such as h-index [29] (e.g. h-Degree [30], and lobby-index [31]) or even hybrid centrality measures [32]. The other sophisticated algorithms are PageRank [21] and HITS [33]. Among them, PageRank [21] is based on Markovian random walk on graphs, which is also considered state-of-the-art in ranking nodes in static networks. Even though these models are based on sophisticated techniques, they do not perform well in dynamic networks [4], [34]. Another way to solve citation prediction by modelling it as a graph is, by counting the predicted links: [35], [36].

The works of Tsoi et al. [37] (Adaptive PageRank and Chang et al. [38] (Lift HITS) consider node labels for first time. These methods are also known as semi-supervised PageRank (SSP) Gao et al. [39]. As these SSPs consider node label, edge attribute as well as node attributes so these are also utilised as objective function for optimization. Further Hsu et al. [40] proposed state-of-the art unsupervised graph ranking model witch also consider node attributes along with network structure. They have considered PageRank and node attributes and proposed random walk based model.

B. NODE RANKING IN DYNAMIC GRAPHS

To consider the temporal aspects of networks in ranking nodes, Newman [41] used re-scaled in-degree of papers in their citation networks, i.e., normalised citations for papers of similar age. A revised equation was proposed by the authors [42], [43]: average numbers of citations for papers published in the same year and in the same field. In a similar quest, it has been discovered that PageRank has a bias toward selecting older nodes as being more important [34]. A re-scaled PageRank model that takes into account moving the temporal window Δt, where Δt is the only parameter to re-scale, has been proposed by Mariani et al. [44]. This model can also be used to adjust the ranking between old and new nodes. When Δt is very small, such that only a few nodes receive links during that time period, it will not be a good predictor and can be considered noise. On the other hand, large *Deltat* leads to ranking older nodes higher. Therefore, it is possible for a hub to have a low degree but still be positioned higher. However, in order to avoid such extreme cases, both situations require careful consideration of the time interval Δt. In a user-item bipartite network, [45] proposed a popularity-based model that combines a node's current degree with its recent degree increase. Meanwhile, in citation graphs, Wang et al. [13] developed variations of PageRank by taking the recent time window Δt into account and incorporating only publications from the last ten years into their model. This consideration of Δt excludes older papers from ranking to avoid the long-term effect they may have on the results. Abbas et al. [46], [47] also proposed models that predict node popularity in temporal bipartite networks through the use of the recent time window and aging effect concepts. Additionally, Long et al. used Eigenvalues over time to rank nodes in dynamic networks [48]. Many real-world data shows heavy tail characteristics [48]

C. DEEP LEARNING ON GRAPHS

Deep neural networks are considered state-of-the-art in many complex learning problems for structured data [49], [50], [51], [52]. Researchers have used deep learning in prediction problems on graphs [53], [54], [55], mainly static graphs. For example [53] have given end-to-end deep-learning based framework for relational reasoning and combinatorial generalization in graphs. Furthermore, [56] proposed a gated-graph neural network for the embedding of different types of nodes, such as START and END. Defferrard et al. [57] presented a generalized convolutional neural-networks for learning and generating local features for graph-like data. This model is a generalization of convolutional neural networks [58] which were originally proposed for image

processing. Node ranking problems can also be modeled as node classification problems. Node classification, which can be a semi-supervised or supervised approach, classifies individual nodes considering the entire network [59]. In another class of node classification problems, node labels are learned from node representations in a vector form [22]: the entire network is first transformed into a low-dimensional representation, and then a node label is learned. However, the node ranking problem has not been explored by the deep-learning community, except for a few researchers who have conducted some work in this direction [60], [61]. These recent research works are only that learn embedding based on the global ranking of the network. Though the difference is that they both are for static graphs. Deep rank considers the attribute of the node along with the network structure. Both works gives unsupervised solution for node ranking problem based on Siamese neural network structure [62]. Siamese networks are good at learning the relationship between pair of objects.

D. CITATION PREDICTORS AND FEATURES

Different features of publications have been used to predict their future citations. For instance, Glänzel and Schubert [63] found the recent citation counts of a paper (e.g., citations received during the past 3–5 years) a reasonable predictor for its future citation counts. Boyack and Klavans [24] considered the importance of the journal it is published in, references, and authors for identifying the importance of papers and found journal importance to be the best predictor, while its correlation value varies by discipline. Chen [64] used structural variation-based metrics (e.g., modularity change rate, cluster linkage, and centrality divergence) in addition to other commonly used metrics such as the number of co-authors, references, and pages (e.g. by Vieira and Gomes [65]). Chen found cluster linkage to be the best predictor of citation counts. Tahamtan et al. [66], in a comprehensive review of the literature, identified 28 factors affecting citation counts that were classified into three main categories: paper, journal, and author-related factors. Bhat et. al. [67] used author interdisciplinarity, author influence, and paper title words along with other generic features such as citation count, age, number of references, authors, etc. Li et al. [68] used peer review text along with abstract text and hand-crafted features for making citation count predictions. Wang et al. [69] presented an extensive analysis of reviews and found review scores to be positively correlated with future paper citations. Li et al. [70] considered citation early trends such as early burst, middle burst, late burst, and so on for making citation count predictions. Ruan et al. [71] solved the citation prediction problem using a back-propagation neural network by considering features such as citations in the first two years, first-cited age, paper length, the month of publication, self-citations of journals, and so on. They have found these features to be more predictive.

E. RESEARCH GAP

The time bias issue is a drawback shared by all of these approaches. Numerous research projects have been carried out in an effort to provide a fair rating of research articles. Some academics concentrate on finding small-scale solutions to time bias issues. For instance, the time-aware PageRank suggested by [72] addresses a number of minor issues with time bias on the relationship between citations of academic papers. According to their findings, eliminating these timing biases can help to increase prediction accuracy. Large-scale temporal biases are the focus of certain other studies [10], [13] while some researchers have addressed the same issue for temporal networks [45], [46], [73]. As it is known paper citation network is always evolving due to the rapid advancement of science and technology, yet mainstream evaluation techniques for research papers are essentially static, making it unfair to compare research publications with various others. For instance, the well-known preferential attachment process results in the typical time bias problem. This issue exists in the science of science at both the scholar and paper levels [7], [74]. Old research publications have an edge in time because they were published earlier than new research articles, which is how this issue may be stated. Therefore, older research papers have had more time to gain attention. Old research articles will be valued higher due to their long-term citation gain, which leads to their eventual global fame. While at the same time, new innovation or discovery happens but due to older paper in the ranking list they are ignored by the peer scientists due to the time biased of algorithms. As a result, it is challenging for new research publications to attract attention. Our findings show that, in addition to the preferential attachment there are other important aspects of previously discussed algorithms that they have not been evaluated on various metrics which is very important for evolving settings. The algorithms previously presented have not been examined on a variety of criteria, which is crucial for circumstances that are always changing. For instance, whatever publications are currently receiving attention or which papers have only just begun to receive attention after being ignored for a considerable amount of time are sometimes referred to as sleeping beauties by researchers [75].

By considering the limitations of previous works, the objective of this study is to propose a learning system that can rank the newly published research papers as well as the research papers that might have been published a long time ago but recently they are gaining attention. While solving the said objective, the system should not ignore the all-time favorite papers either. To achieve this, we have evaluated the proposed system on 6 evaluation metrics.

III. METHODS

This study aims to predict the popularity of academic publications based on their citation counts, which serve as an indicator of future citations. To achieve this goal, the

study focuses on directed temporal networks, which consist of citing papers (o_i) and cited papers (o_j) belonging to the sets V. A direct link is formed when paper o_i cites paper o_j. The formed graph or network can be represented as an adjacency matrix (A) where each entry (i, j) indicates whether a link exists between a citing and cited paper. The in-degree (k_o), i.e., the number of papers citing a paper o, is used to measure its citation count. The cumulative citation count of a paper o at any time t is calculated using the time-based adjacency matrix $A(t)$ and serves to predict its future citation counts. The formula for the total citation count of a paper o at any time t is as:

$$k_o(t) = \sum_u A_{uo}(t) \tag{1}$$

$$\Delta k_o = k_o\left(t_f\right) - k_o(t) \tag{2}$$

Prior to presenting our suggested citation prediction models, we provide a brief overview of three fundamental benchmark models. Firstly, the well-established PageRank algorithm is used for ranking in static graphs. Secondly, we introduce the Support Vector Regressor (SVR) machine learning model, which is frequently used in citation prediction. Finally, we discuss the Temporal-Based Model (TBP), which is designed for ranking in temporal networks.

A. PageRank

PageRank was originally developed as a ranking algorithm for web pages within the Google search engine [21]. However, its applications can extend to ranking nodes in other networked systems where the structural attributes of the nodes play a significant role, such as in the ranking of scientific papers and authors [76]. The process of the PageRank algorithm is described as follows: If node o_i has a link to node o_j, a directed link is created between them ($o_i - > o_j$). If a node (e.g., a web page or paper) possesses a set of links to other nodes, denoted as S_i, its significance will be distributed among the nodes in set $|l_j|$, where $|S_i|$ represents the number of nodes. The transition matrix for the network or graph A can be expressed in the following manner:

$$A_{ij} = \left\{ \begin{array}{ll} 1/l_j \; if \; n_j \in s_i \\ 0 \quad Otherwise \end{array} \right\} \tag{3}$$

According to the PageRank algorithm, papers that are cited by highly cited papers have a greater likelihood of attracting more citations. However, the presence of unconnected nodes, or *dangling nodes*, can complicate the analysis. To address this issue, a transformed matrix (S: $S = A + N_{cd}$) can be introduced.

With the exception of the column identified as 'dangling nodes', denoted as $1/N$, all components of the matrix Ncd are zero. Here, N represents the number of rows or nodes in the matrix. These columns undergo normalization such that their summation equals 1, producing a column stochastic matrix. Consequently, the PageRank for 'dangling nodes' will not be zero. Assuming a user follows the PageRank (as directed by matrix S) with probability (α), there is a probability of $(1 - \alpha)$ that the user will randomly select a page during navigation. Hence, the expression for PageRank can be formulated as follows: $M = \alpha S + \frac{(1-\alpha)}{n} I_n$. The matrix I_n denotes an $n \times n$ square matrix consisting of n elements. Using the power method, one can compute the PageRank vector (PR), where if $PR^k = M.PR^{(k-1)}$, it converges to a stationary vector referred to as PageRank.

1) TEMPORAL-BASED PREDICTOR (TBP)

In the temporal-based model developed by [73], the calculation of a node's ranking score for future link gain takes into account the effects of link decay. This consideration of link decay's impact is an important feature of the model.

$$s_o(t) = \sum_u A(t) \exp(\gamma(T_{o'o} - t)), \tag{4}$$

where the model factors in the prediction score, denoted as $s_o(t)$, which pertains to the object of interest at a given time t. The user-object (in this case, paper-paper citation) adjacency matrix at time t, referred to as $A(t)$, is also taken into consideration. Furthermore, the time at which the node o received a link from other nodes, which denotes the creation time of a related node o', is marked as $T_{o'o}$. It is important to note that this event occurred before the time of interest t, meaning $T_{o'o} < t$. Additionally, the link decay rate, denoted as γ, is incorporated into the model's predictions.

2) SUPPORT VECTOR REGRESSOR (*SVR*)

As another baseline, we use the Support Vector Regressor model with RBF kernel [77]. For *SVR* learning, we use similar inputs and outputs for training as in our proposed model.

B. OUR PROPOSED DEEP LEARNING FRAMEWORK

Deep learning is part of the machine learning family of methods, which are the state-of-the-art patterns learning algorithms [49], [51]. It mimics biological neural-networks, which are also known as artificial neural networks. There are many architectures to learn the pattern from different kinds of problems, such as deep neural networks, convolutional networks, recurrent neural networks and so on. We are using deep neural networks, which have many (two in our case) layers between the input and predicted output layers. These kinds of connected networks are also known as "fully connected neural networks," which can be divided into three main parts: 1) input layers; 2) hidden layers; and 3) output layers. Input layers are the first layers that take the example data as input to be learned. Hidden layers are intermediate layers between the input and output layers. The output layer is the final layer, which gives the predicted result.

To solve the learning problem, we take input and at every node in each layer apply some transformations, which are then taken as input to the next layer, and so on as we reach the final output layer. To learn the parameter values, we take the gradient at the final layer and apply the chain rule in the reverse direction up to the initial layer to find gradients of each intermediate layer's parameters. As the output of every

node is taken as input in the next layer, it is possible to learn non-linear hypotheses very easily. It is noteworthy to highlight that all states of a graph before time t are inputs to our model, and the predicted ranking scores are the model's outputs.

Our proposed framework is shown in Algo 1. So we have learned the ranking problem as follows:

1) INPUT FEATURES

We extract the following three features from temporal graph snapshots:

1) *Node degree:* As already explained, node degree $k_o(t)$ at time t can be easily calculated just by counting the total number of links that a node has at time t.
2) *Aging effect:* To consider the ageing effect at any time t, as $a_o(t) = exp(t - t_b)$, where t_b is the birth time of the node in the system.
3) *Total influence with varying decay:* We consider that every new link to a node (paper) influences the future possibility of new links. i.e., its influence decays exponentially over time. As a paper gets more links, its influence decay rate decreases. To consider this effect, we consider the decay rate inversely proportional to the current degree of the paper. i.e. $\sum_{o}^{'} A(t)\exp(\frac{(T_{o'o}-t)}{k_o(t)})$, where $T_{o'o}$ is the time when node o has received link from other nodes. In other words, if a node's current degree is higher, its influence decay rate will be lower, increasing the probability of attracting new links. Conversely, if a node's current degree is low, then its decay rate will be higher, which ends up having a low probability of attracting new links.

2) PARAMETER INITIALIZATION

For parameter initialization (W) at each layer, we generated random numbers from an uniform distribution, i.e., $U \in [0, 1]$. Then we divide each random number by $\sqrt{\frac{layerIndegree+layerOutDegree}{2}}$ i.e., the average of the total in-degree of the layer l and the total out-degree of the layer l. Parameter b is simply initialised with zero.

3) FORWARD PROROGATION

In forward prorogation, we used Tanh (hyperbolic tangent) activation, which we found gave a slightly better result than Relu activation in our case. In the final layer, we used Sigmoid activation,see table [1] for detailed information. We applied the following steps to calculate the activation at layers 1, 2 and 3. Let the input feature matrix be X, and randomly initialise parameters $W^{[l]}$ and $b^{[l]}$ at every layer l. The forward propagation step is at layer 1 ($l = 1$).

$$Z^{[1]} = W^{[1]}.X + b^{[1]} \tag{5}$$

$$a^{[1]} = g^{[1]}(Z^{[1]}) \tag{6}$$

$$Z^{[2]} = W^{[2]} * a^{[1]} + b^{[2]} \tag{7}$$

where X is the input feature vector, $W^{[1]}$ and $b^{[1]}$ are parameters to learn, which are randomly initialised as stated in section [III-B2]. $Z^{[1]}$ is used to calculate activation $a^{[1]}$ at layer 1. From e.q.[5] to e.q.[7], we can see how to propagate from layer 1 to 2 to calculate activation $a^{[2]}$. We repeat this process two more times to calculate the activation function for layers 2 and 3. Meanwhile, we also cache those values to calculate gradients in backward prorogation.

At intermediate layers, the $tanh(z) = (e^{2z} - 1)/e^{2z} + 1$ activation is used, and at the final layer, the $sigmoid(z) = \frac{1}{(1+e^{-(z)})}$ activation is used.

4) ACTIVATION FUNCTIONS

Activation functions play a crucial role in neural networks because, during the flow of data from input to output layers, they transform the data from the current layer into an output using some non-linear functions. Because of activation functions, neural networks are gaining more success as they turn linear classifiers into non-linear ones. The activation functions behave differently for different problems, therefore, so far, many activation functions have been proposed [78]. The adaptation of activation functions over different layers depends on the choice of the problem too. For example, the sigmoid and hyperbolic tangent (tanh) at the final layer are generally used for binary classification because it crunch the number between [0, 1] and [−1, 1] respectively. Linear or identity functions are used for regression, and Softmax is used for multi-class classification and so on, though they are not restricted only to these problems. There are more choices for activation functions in hidden layers. For example, the Sigmoid activation function in hidden layers was considered the best until the new activation Relu was discovered, which is considered the best for many problems [79]. So in intermediate layers, many activation functions can be used according to the problem under consideration [78] or sometimes even search for the best activation functions [80]. Our problem is a bit different from the classification or regression problem. Therefore, we converted our problem into a probabilistic regression problem, and used Sigmoid activation at the final layer, which outputs in [0, 1]. Probabilistic regression is used to make our model suitable for learning purposes, i.e., to make the parameter learning better.

5) COST FUNCTION

We converted our ranking problem to a regression problem by converting the node's ranking score into a normalised score. So the ranking score for supervision is normalised as follows:

$$\Delta k_o(t, T_F) = k_o(T_F) - k_o(t) \tag{8}$$

$$Y_o = \frac{\Delta k_o(t, T_F)}{\sum_{u} \Delta k_{o'}(t, T_F)} \tag{9}$$

$$Y_{o\_normalized} = \sum_{Y_{o'} <= Y_o} Y_{o'} \tag{10}$$

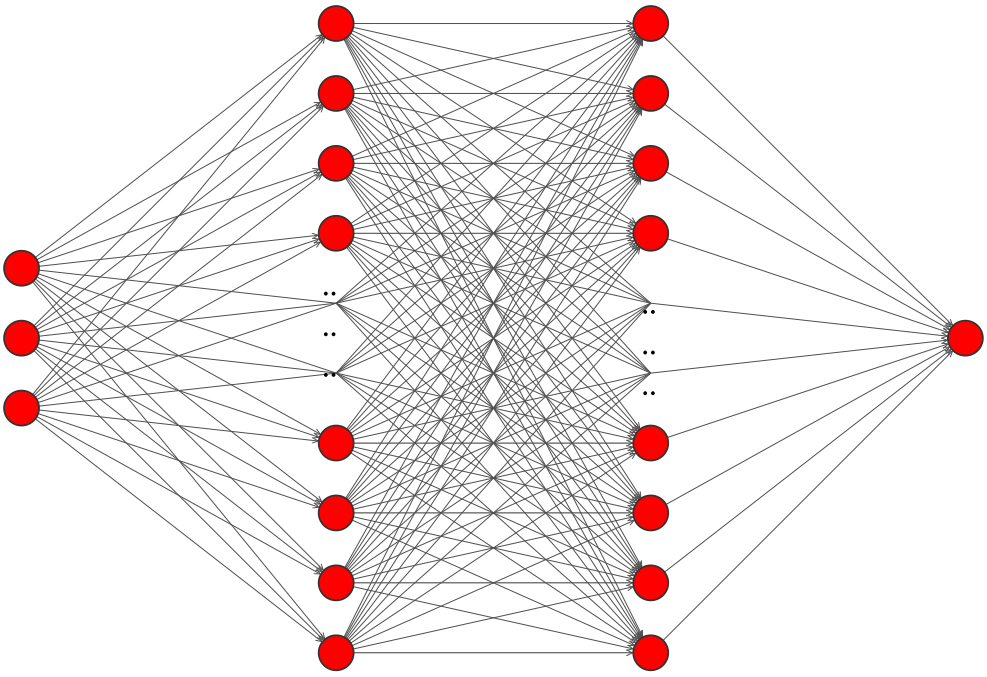

FIGURE 1. This is architecture of our fully connected neural network. The dimensions of the hidden layer parameters Ws are (3 × 50) and (50 × 1) respectively. Consequently parameter b is re-scaled to match with W.

TABLE 1. Parameters used in our DNN architecture.

Layer	Number of Neurons	Activation Function	Description
Input layer	3	**tanh** $\in (-1, 1)$	After feature extraction operation it is fed to DNN
Hidden layer	50	**tanh** $\in (-1, 1)$	First hidden layer
Hidden layer	50	**tanh** $\in (-1, 1)$	Second hidden layer
Output	1	**sigmoid** $\in [0, 1]$	This layer output the ranking score

We use $Y_{o\_normalized}$ for supervision while parameter learning. Normalizing in this way makes sigmoid activation a suitable choice. As $max(Y_{o\_normalized})$ will give the score 1 and also $Y_{o\_normalized} \in [0, 1]$. Furthermore, we also normalise the features, except for the ageing score feature, in a similar manner and calculate the cost as cross entropy loss, which is also described as point-wise loss in ranking problems [81].

$$Cost(Y_{o\_normalized}, \widehat{Y}_o) = -Y_{o\_normalized} \log(\widehat{Y}_o) \cdots - (1 - Y_{o\_normalized}) \log(1 - \widehat{Y}_o) \tag{11}$$

6) BACK PROROGATION

We performed the following actions while performing back propagation to determine gradients:

$$\frac{\partial Cost(Y_{o\_normalized}, \widehat{Y}_o)}{\partial \widehat{Y}_o} = d\widehat{Y}_o = da^{[3]} = \cdots - \frac{Y_{o\_normalized}}{\widehat{Y}_o} + \left(\frac{1 - Y_{o\_normalized}}{1 - \widehat{Y}_o}\right) \tag{12}$$

We back propagate three levels using the chain rule to discover the following derivatives:

$$dZ^{[3]} = da^{[3]} * g^{[3]'}(Z^{[3]}) \tag{13}$$

$$dW^{[3]} = dZ^{[3]}.a^{[2]} \tag{14}$$

$$db^{[3]} = dZ^{[3]} \tag{15}$$

$$da^{[2]} = W^{[3]^T}.dZ^{[3]} \tag{16}$$

where $dZ^{[3]}$,$dW^{[3]}$,$db^{[3]}$ and $da^{[3]}$ are derivatives of loss function with respect to Z, W, b and a at layer 3. As we can see from e.q. [13] to e.q. [16] using $da^{[3]}$ as input, we find $da^{[2]}$. We proceed to repeat the aforementioned process twice more in order to estimate gradients of parameters in our 1st layer. Furthermore, it is possible to express the derivatives of the activation function as follows:

$$sigmoid(z)' = sigmoid(z) * (1 - sigmoid(z)) \tag{17}$$

$$\tanh(z)' = (1 - \tanh^2(z)) \tag{18}$$

Algorithm 1 The Pseudo-Code of the Proposed Rank Aware Learning Framework for Dynamic Citation Graph

Require: $G(V, E_T)$
Ensure: $T > 0, E_T \notin \phi$
Step 1: Input: From citation data convert to graph $G(V, E_T)$.
Step 2: Extract node level temporal features from graph snapshots($G_1, G_2 \ldots, G_T$).
Step 3: Create ground truth vector $Y_{o\_normalized}$ which is suitable for our ranking based loss function(See III-B5).
Step 4: Learn d-dimensional node embedding.
Step 5: Repeat step-3 and step-4 for 5 random time points.
Step 6: return (a ensembled classifier based on above 5-trained models.)

C. DATA SET

The precision and efficacy of the predictors in our proposed model were examined through experimentation with the arXiv paper citation dataset. This dataset consists of citation information of papers indexed in the arXiv database of High Energy Physics. In the following paragraphs, we present a more comprehensive overview of the pre-processed data set, which is also publicly available [82]. Time binnings were achieved through pre-processing, averaging at monthly intervals. Figure[2] displays in-depth statistics of the dataset. The arXiv-HePh dataset contains 30,500 papers and 347,185 citation connections, spanning from January 1993 to April 2003.

IV. EXPERIMENTAL SIMULATION, RESULT ANALYSIS, AND COMPARISON

In this section, we will begin by outlining the random sampling strategy that was employed to acquire our data. Additionally, we will explore our optimization methodology, which involved utilizing the Adam optimizer alongside its corresponding training parameters for our deep-learning algorithm. Subsequently, we will provide a succinct overview of our ensemble-learning strategy, which was implemented with the aim of enhancing the overall predictive capabilities of our models. Following this, we will present four distinct evaluation metrics, namely Temporal Novelty (TN_k), Precision, Novelty, Area Under Receiving Operating Characteristic (AUC_k), which is commonly referred to as ROC, Kendal's rank correlation Tau (τ), and Normalized Discounted Cumulative Gain ($NDCG$). These specific metrics were utilized to accurately measure the overall performance of our models. Finally, we will display and analyze the results of our assessment of varying scenarios. These include instances where future time (T_F) varies while past time (T_P) remains consistent, in addition to scenarios where both past and future times (T_F and T_P, respectively) vary while the size of k remains fixed.

A. SAMPLING DATA FOR EXPERIMENTS

In our study, we follow a random time-based approach to evaluate the performance of our model. To ensure adequate evaluation, we consider a considerable size of past and future time windows. Accordingly, we rank the nodes at the randomly selected time t and evaluate the efficacy of our model at the subsequent time point, $t + T_F$. Our training dataset comprises interaction data up to time t, while the test dataset comprises interaction data taken between t and $t+T_F$.

B. PARAMETER OPTIMIZATION FOR TRAINING DEEP-LEARNING

We employed the Adam optimizer [83] in our parameter learning process as it has been demonstrated to locate optimal minima more efficiently than Gradient descent. This optimization technique involves calculating an exponentially weighted average of past gradients and correcting for biases ($v^{corrected}$). Additionally, it computes an exponentially weighted average of the squares of past gradients and adjusts for biases ($s^{corrected}$). The specific equations used are presented below:

$$v_{dW^{[l]}} = \beta_1 v_{dW^{[l]}} + (1-\beta_1) dW^{[l]} v^{corrected}_{dW^{[l]}} = \frac{v_{dW^{[l]}}}{1-(\beta_1)^t} \tag{19}$$

$$s_{dW^{[l]}} = \beta_2 s_{dW^{[l]}} + (1-\beta_2)(dW^{[l]})^2 s^{corrected}_{dW^{[l]}} = \frac{s_{dW^{[l]}}}{1-(\beta_1)^t} \tag{20}$$

$$W^{[l]} = W^{[l]} - \alpha \frac{v^{corrected}_{dW^{[l]}}}{\sqrt{s^{corrected}_{dW^{[l]}}} + \epsilon_d} \tag{21}$$

$$b^{[l]} = b^{[l]} - \alpha \frac{v^{corrected}_{db^{[l]}}}{\sqrt{s^{corrected}_{db^{[l]}}} + \epsilon_b} \tag{22}$$

In these equations, t indicates the number of steps taken by Adam, l represents the number of layers $(1, 2, 3)$, β_1 and β_2 are hyper-parameters that regulate the two exponentially weighted averages. Additionally, α denotes the learning rate and ϵ is a very small number included to prevent division by zero. Notably, parameter $b^{[l]}$ is also updated for each layer l using the same aforementioned steps. In our implementation, we maintained the learning rates at 0.0075, $\beta_1 = 0.9$, $\beta_2 = 0.99$, $t = 2$ and $\epsilon = 1e-8$. We executed 10, 000 iterations over these parameters.

C. ENSEMBLE LEARNING

If we learn the parameter at one specific random moment, it might not be able to generalize across another time step t because our problem is time-varying. To solve this issue, we picked 5 random locations inside our temporal graph

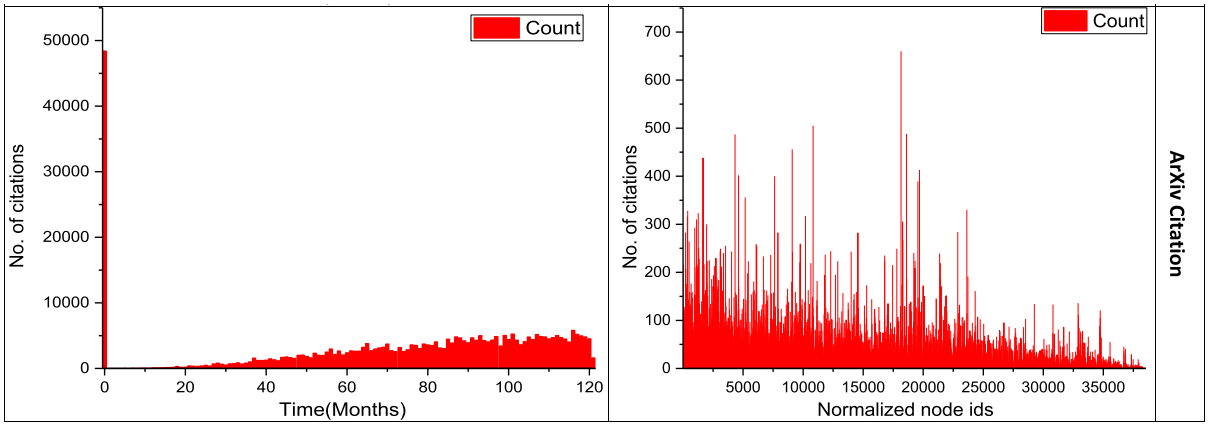

FIGURE 2. The statistics of the data used in our experiments are shown in the accompanying figure. The number of citations per unit of time is depicted in the left figure. The number of citations for each paper, or node in our case, is shown in the right figure.

timeline and determined the parameter by taking into account that $T_P = T_F = 10$ months. Finally, we simply averaged over 5 anticipated activation ratings and ordered them when generating predictions. The good news is that even though we only considered $T_P = T_F = 10$ when learning the parameters, we were still able to predict up to 40 months.

D. EVALUATION METRICS

We have adopted the following six evaluation metrics to measure the accuracy of our proposed model:

1) *Temporal Novelty(TN_k)* assesses the predictive capacity of a model by measuring its ability to accurately rank 'new nodes' that have recently gained popularity but were overlooked during an earlier time frame. Specifically, for a future time window T_F corresponding to the interval during which these nodes accrued popularity, it quantifies how well the model can identify these nodes among the top k rankings. Assuming $R_k^{\Delta t}$ defines the number of new objects that are not present among the top rankers based on their popularity gain during the recent past time window T_P and $E_k^{\Delta t}$ denotes the number of correctly predicted new objects in the top k rank by our model, we can calculate the temporal novelty (TN_k) score using the following formula:

$$TN_k = \frac{E_k^{\Delta t}}{R_k^{\Delta t}}, \tag{23}$$

2) *AUC*, or "ROC," which was introduced by Hanley and McNeil [84], provides a measure of the importance of the relative position of the top k items in both a predicted and a ranked list. Specifically, this metric selects the top k items from a real list and uses their rank scores as a benchmark for comparison with the top k items in a predicted list. To calculate the AUC, one can consider the scores of an object in the predicted list $s_p \in L_p$ and in the real list $s_r \in L_r$, and apply the following formula:

$$AUC = \frac{\sum_{s_p \in L_p} \sum_{s_r \in L_r} I(s_p, s_r)}{|L_p|\,|L_r|} \quad where, \tag{24}$$

$$I(s_p, s_r) = \begin{cases} 0, & \text{if } s_p > s_r, \\ 0.5, & \text{if } s_p = s_r, \\ 1, & \text{if } s_p < s_r. \end{cases} \tag{25}$$

3) *Precision:* The definition of *precision* is the ratio of the number of objects present in the top k rankings of the predicted and actual ranking lists [85], and is expressed as:

$$P_k = \frac{D_k}{k}, \tag{26}$$

In this equation, D_k represents the count of common objects that appear in the top k positions of both the predicted and actual ranking lists. The value of $P_k \in [0, 1]$, where a higher value of P_k indicates a more accurate prediction.

4) *Novelty(Q_k)* evaluates how effectively a predictor can rank 'new objects' in the top k positions which were not in the top k positions previously. Let R_k denote the count of 'new objects' present in the top k positions of the actual list, and E_K represent the number of 'new objects' correctly predicted by our model in the top k positions of the ranking list. The novelty score is computed as follows:

$$Q_k = \frac{E_k}{R_k}, \tag{27}$$

5) *Kendall's Tau (τ)* quantifies the correlation between predicted and actual rankings, with values ranging from -1 to $+1$. A value of $\tau = 1$ indicates perfect agreement between the predicted and actual rankings, $\tau = 0$ suggests independence between the rankings, and $\tau = -1$ indicates complete disagreement between

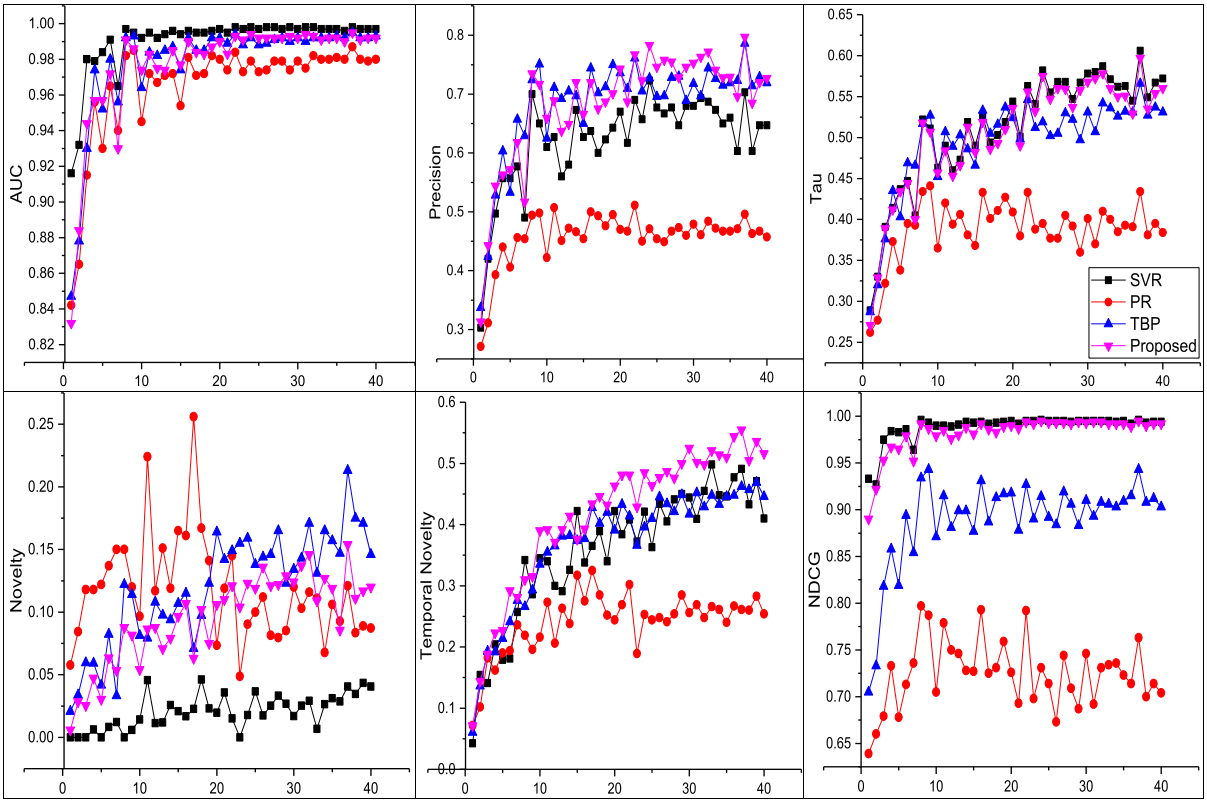

FIGURE 3. The sensitivity of the model's performance across various future time window sizes (X-axis) is demonstrated in the figure above. The past time window is kept fixed at 10 months. The Y-axis represents the accuracy score, with higher scores indicating better performance. All values fall within the range of 0 to 1, except for the rank correlation tau (τ), which ranges from -1 to 1. The solid black line with a square symbol represents the performance of the Support Vector Regression (SVR) model. The solid blue line with a triangle symbol is indicative of the TBP model, while the solid red line with a circle symbol shows the performance of the PageRank (PR) model. The proposed model's performance is represented by the pink line with a downward triangle symbol.

them. Kendall's Tau can be expressed as:

$$\tau = \frac{C - D}{\sqrt{(C + D - N_{tp})}\sqrt{(C + D - N_{tr})}}, \tag{28}$$

where, the number of concordant pairs, denoted by C, and the number of discordant pairs, denoted by D. Additionally, N_{tp} is the number of ties in the predicted list, and N_{tr} represents the number of ties in the real list.

6) *Normalized Discount Cumulative Gain (NDCG)*, which was proposed by [86], involves evaluating the performance of the top k objects in a ranked list. If r_i represents the relevance level of the o_i objects within the top k, then the *NDCG* score can be computed as follows:

$$NDCG_k = \frac{r_1 + \sum_{j=2}^{k} \frac{r_j}{\log_2(1+\Pi(o_j))} I(\Pi(o_j), k)}{r_1 + \sum_{j=2}^{k} \frac{r_j}{\log_2(1+g(o_j))} I(g(o_j), k)} where, \tag{29}$$

$$I(x, y) = \begin{cases} 1, & \text{if } x <= y, \\ 0, & otherwise. \end{cases} \tag{30}$$

Here, $g(o)$ and $\Pi(o)$ indicate the ranking position of object o in the actual and predicted lists, respectively.

E. RESULTS ANALYSIS AND COMPARISON

1) VARYING FUTURE TIME WINDOW (T_F), FIXED PAST TIME WINDOW (T_P) AND FIXED SIZE k

In Fig.3 we have shown the performance of our model when only the future time window is varying. To test some of the important evaluation metrics such as novelty and temporal novelty, we set a fixed past time window ($T_P = 10$) and vary the future time window T_F up to 40 months. The X-axis shows the time, and the Y-axis shows the accuracy results based on different evaluation metrics. Higher results depict better performance. All the evaluations are considered by considering top 100 nodes except for rank correlation (τ) which considers the total number of items in the list. Please note that after randomly selecting a random training point at time t, we rank the nodes according to citation gain during a time window T_F after t. Our model makes use of all the information before time t to calculate the ranking scores for each node at time $(t + T_F)$, and determines the top k scored nodes during each time window. From Fig.3 following analysis can be generated.

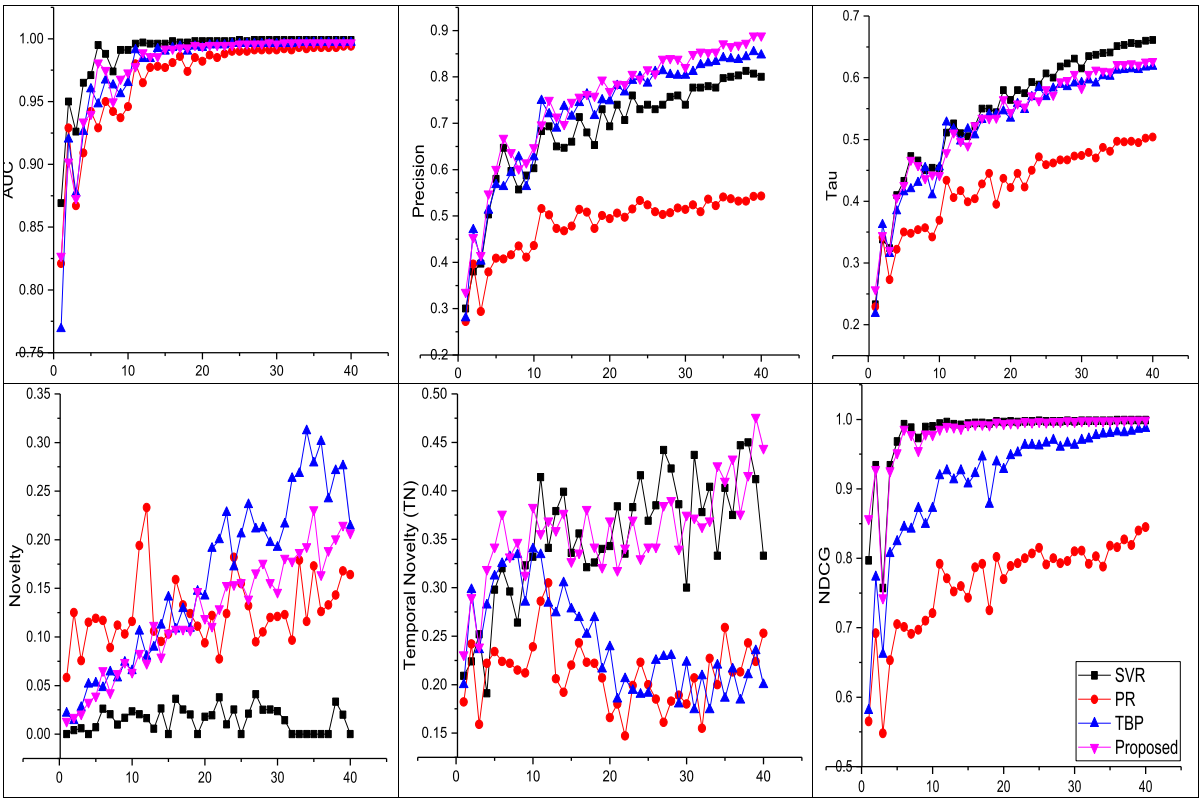

FIGURE 4. The figure displays the sensitivity of the proposed models when both the T_F (future time window) and T_P (past time window) are varied from a period of 1 to 40 months. At each data point, the past and future time windows are kept equal, i.e., $T_F = T_P$. The solid black line with a square symbol represents the performance of the Support Vector Regression (SVR) model. The solid blue line with a triangle symbol is indicative of the performance of the TBP model, while the solid red line with a circle symbol shows the performance of the PageRank (PR) model. The pink line with a downward triangle symbol represents the performance of our proposed model.

1) In *AUC* analysis, it is found that our proposed model performs the same as TBP while PageRank (PR) underperforms. It is also noticed that SVR outperforms all by a slight margin.
2) In *Precision* analysis, we have found the same behavior as AUC analysis. But PageRank underperforms with a higher margin. It is also found that as the time window increases our proposed model seems to dominate all, while considering AUC, almost all methods show similar results for a higher time window (more than 20).
3) In *Tau* (τ) rank correlation analysis, our proposed model and SVR show a slightly better performance after 25 month. Initially, TBP and PageRank have similar performance but later TBP outperforms PageRank.
4) In *Novelty* analysis, our proposed model is less fluctuated compared to the other three methods and has the lowest initial value, which improves compared to PageRank and SVR but is still not as high as TBP. TBP is consistently increasing with time, while PageRank initially shows better performance but after some time, its performance decreases. SVR shows the worst performance on predicting novel entries in the top 100 ranking list.
5) In *temporal novelty (TN)* analysis, we find our proposed model consistently performs better. The second performer is TBP and then SVR and then PageRank.
6) In *NDCG* analysis, our proposed model and SVR outperform consistently with a high margin, while TBP performs better than PageRank.

2) VARYING FUTURE (T_F) AND PAST (T_P) TIME WINDOWS AND FIXED k SIZE

In Fig. 4, we have shown the performance of our model when both future and past time window sizes are varying (at every time step $T_P = T_F$) while k size is fixed.

1) In *AUC* analysis, all four predictors show a very similar performance, particularly for time windows larger than 15.
2) In *precision* analysis, almost all models perform consistently increasing with the size of the time window. Nevertheless, our proposed model outperforms others followed by TBP, SVR, and PageRank.
3) In rank correlation *Tau* (τ) analysis, our proposed model, SVR and TBP perform very similar. However, after a certain time SVR outperforms the others. PageRank is found to be the least performer.

4) In *novelty* analysis, initially PageRank was the best performer but after sometime its performance remain almost similar while fluctuates. While the performance of TBP and our proposed model consistently increases, after sometimes TBP outperforms the proposed model. SVR is the worst-performing model.
5) In *temporal novelty (TN)* analysis, our proposed model and SVR consistently outperforms all the predictors with high margin. The third performer is TBP, with a decreasing trend, and then PageRank with almost similar with a lot of fluctuations.
6) In *NDCG* analysis, our proposed model and SVR outperform consistently with a high margin. Although TBP's performance was lower but it almost reaches the other close to time window size 40 and PageRank again is the least performer with a high margin while having an increasing trend.

V. DISCUSSION

Predicting the future impact of scientific publications and consequently authors, institutes, and even countries are important but challenging task. The sole use of quantitative metrics such as citations for measuring and comparing scientists, institutes, and countries are highly adopted because it is the most widely and easy-to-use method. This research aims to predict scientific impact considering its temporal aspects taking into account its decay over time. The impact of publications can be found by modeling publications' citation interactions in a temporal citation network and ranking the publications as the nodes in the network. There are many sophisticated models for the ranking problem such as PageRank that consider the whole network into account. But these models have two main problems: not considering the temporal effect, and their complexity increases as the network size increases, or as new nodes enter the system it needs to be re-calculated. Therefore, this work presents a deep neural-network-based framework, which only considers local node-level temporal information in ranking nodes globally. We extracted three simple temporal network features by considering nodes' local temporal information such as current total degree, birth time, and the time they receive new links. It is crucial but difficult to predict the future impact of scientific papers and, by extension, authors, institutes, and even nations. The most popular and straightforward technique for measuring and comparing scientists, institutions, and nations is the exclusive use of quantitative indicators like citations. This study seeks to forecast scientific effect by taking into consideration its temporal components and accounting for their deterioration with time. Modeling the citation interactions of publications in a network of temporal citations and rating the publications as the network's nodes allows one to determine the influence of publications. Our proposed model is not always the best, but on average it is the best considering all six evaluation metrics. The model's performance can be improved by providing additional time and data, as in the case of the citation, 20 years are not enough to see all the effects in the system.

VI. CONCLUSION, LIMITATIONS, AND FUTURE WORKS

A delicate part of scientific evaluation jobs is citation time. Therefore, for our analysis in this study, we used 20 years' worth of citation data. We also tested the robustness of our proposed model using six common information retrieval-based measures. Our architecture is straightforward and easily adaptable to other temporal networks. Figures [3 and 4] from our experimental investigation on citation data demonstrate that, on average, our suggested framework outperforms the other three baseline techniques, namely PageRank, SVR, and TBP. According to the data, our suggested model performs significantly better than the other three models for NDCG and TN assessment metrics than all other accuracy metrics combined. The presented deep-learning-based framework can be easily applied to other node-ranking problems in temporal networks.

Ranking nodes in dynamic or temporal networks can be useful in a variety of applications. Some examples include:

1) Identifying influential nodes in social networks: Ranking nodes in social networks can help identify influential individuals or organizations based on their connections and influence on the network over time.
2) Detecting anomalous behavior in network data: Ranking nodes in dynamic networks can help detect anomalies or unusual behavior by comparing the rankings of nodes at different points in time.
3) Understanding network evolution: ranking nodes in temporal networks can provide insight into how the network evolves over time and how the roles of different nodes change.
4) Predicting future behavior: Ranking nodes in dynamic networks can be used to make predictions about future behavior based on past patterns.

Overall, ranking nodes in dynamic or temporal networks can provide valuable information about the structure and evolution of the network, as well as identify important or anomalous nodes and behavior.

The proposed model's capacity to generalize is good because it learns parameters at chosen time points at random. Since it is built on an ensemble scoring mechanism, the model may be trained at various points in time and used to produce predictions for large, continuously evolving networks. Therefore, it is a good choice for implementing in real-time scenarios. One of the drawbacks of our suggested findings is that explicit node-level features must be concatenated with the existing framework before being passed to the DNN. We have not used the explicit features of nodes in the current dataset setup, hence we are unable to provide quantitative feedback on such instances.

Some limitations of our model include:

1) **Data requirements:** The proposed framework often requires large amounts of data in order to generate

TABLE 2. List of Abbreviations and symbols.

Abbreviation	Definition
G	Graph
V	Set of nodes
o	A node from set V
E	Set of edges
A	Adjacency matrix
A(t)	Adjacency matrix at any time t.
E_T	Set of temporal edges between any two node i and j.
t	The time at which a paper has received citation
Δt	Duration of time window
$k_o(t)$	Total degree of node o at any time t.
Δk_o	Degree gain by node o during any time duration Δ t
T_f	Time at any future time point f.
PageRank/PR	A ranking algorithm based on Markovian random walk.
SVR	Support Vector Regressor, one of the benchmark model.
TBP	Temporal Based Predictor, one of the benchmark model.
so(t)	Ranking score my model at any time t for node o.
γ	Decay rate
α	Random walk transition probability
S	Intermediate matrix
N_{cd}	Colulmn normalized intermediate matrix of size A.
In	Identity matrix
M	Intermediate matrix some times called Google Matrix
$a_o(t)$	Feature for node o at time t.
X	Input feature matrix of Neural Network
W	Is parameter used in Neural network
DNN	Deep Neural Network
tanh(z)	Hyperbolic tangent function used as activation function.
Z	Learned value after each layer of NN.
g()	Any Activation function
$\Delta\ k_o(t, T_F)$	Degree gain by node o between time t and any future time TF.
Yo	Is normalized degree gain for all nodes.
Yo_normalized	Is Cumulative Distribution Function for rank score of node o.
sigmoid(z)	Sigmoid function
dZ	Partial derivative of function Z.
$\beta 1$ and $\beta 2$	Hyper-parameters in adam optimizer
ϵ	Very small value
L,l	Layer of Neural network
TN_k	Temporal Novelty evaluation metrics for top k node.
AUC_k	Evaluation metrics: area under the ROC curve (0,1)
P_k	Evaluation metrics: Precision of top k items (0,1)
Q_k	Evaluation metrics: Novelty score of top k items (0,1)
τ	Evaluation metrics: Kendal's Rank correlation score (-1,1)
NDCG k	Evaluation metrics: Normalized Discount Cumulative Gain of top k items (0,1)
C	number of concordant pairs in rank correlation calculations.
D	number of discordant pairs in rank correlation calculations.
N_{tp}	Number of ties in the predicted list in rank correlation calculations.
N_{tr}	Number of ties in the in rank correlation calculations.

accurate results. This can be a challenge in situations where data is scarce or difficult to obtain.

2) **Assumptions:** The proposed framework often rely on certain assumptions about the data and the underlying

process being modeled. These assumptions may not always hold, which can lead to inaccurate or misleading results.

3) **Limited interpretability:** The proposed framework may be less interpretable than other approaches such as statistical models, making it difficult to understand the factors that are driving the rankings.
4) **Complexity:** The proposed framework can be complex, requiring the use of advanced statistical or machine learning techniques. This can make them difficult to implement and interpret for some users.
5) **Bias in the data:** The proposed framework may be subject to biases in the data that it is trained on, which can lead to biased or unfair rankings.

Overall, while the proposed framework can provide powerful tools for analyzing temporal networks, it also comes with a number of limitations that should be taken into consideration.

In future work, one can approach the ranking problem while considering the above-mentioned problems. There are many potential areas of focus for future work on DNN-based ranking problems. Some potential areas of focus include:

1) **Developing more robust and accurate ranking models:** Researchers could work on developing DNN-based ranking models that are more robust and accurate, particularly in situations where data is scarce or noisy.
2) **Incorporating more diverse data sources:** Many ranking problems involve data from multiple sources, such as text, images, and audio. Future work could focus on developing DNN-based ranking methods that can effectively incorporate and make use of data from multiple sources.
3) **Improving interpretability:** DNN-based ranking methods can be difficult to interpret, making it challenging to understand the factors that are driving the rankings. Future work could focus on developing methods that provide more interpretability, such as using explainable AI techniques.
4) **Addressing bias:** DNN-based ranking methods may be subject to biases in the data that they are trained on, which can lead to biased or unfair rankings. Future work could focus on developing methods that are more resistant to bias and that can produce fairer rankings.

Overall, there are many directions that future work in DNN-based ranking problems could take, including improving accuracy and robustness, exploring different ranking approaches, incorporating diverse data sources, and addressing bias.

APPENDICES

A. COMMONLY USED NOTATIONS

We have shown the abbreviations and notations in the following table 2.

REFERENCES

[1] G. Iñiguez, C. Pineda, C. Gershenson, and A.-L. Barabási, "Dynamics of ranking," *Nature Commun.*, vol. 13, no. 1, p. 1646, Mar. 2022.
[2] M. K. Hasan, T. M. Ghazal, A. Alkhalifah, K. A. A. Bakar, A. Omidvar, N. S. Nafi, and J. I. Agbinya, "Fischer linear discrimination and quadratic discrimination analysis–based data mining technique for Internet of Things framework for healthcare," *Frontiers Public Health*, vol. 9, p. 1354, Oct. 2021.
[3] M. K. Hasan, S. Islam, I. Memon, A. F. Ismail, S. Abdullah, A. K. Budati, and N. S. Nafi, "A novel resource oriented DMA framework for Internet of Medical Things devices in 5G network," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 8895–8904, Dec. 2022.
[4] H. Liao, M. S. Mariani, M. Medo, Y.-C. Zhang, and M.-Y. Zhou, "Ranking in evolving complex networks," *Phys. Rep.*, vol. 689, pp. 1–54, May 2017.
[5] M. K. Hasan, M. Akhtaruzzaman, S. R. Kabir, T. R. Gadekallu, S. Islam, P. Magalingam, R. Hassan, M. Alazab, and M. A. Alazab, "Evolution of industry and blockchain era: Monitoring price hike and corruption using BIoT for smart government and industry 4.0," *IEEE Trans. Ind. Informat.*, vol. 18, no. 12, pp. 9153–9161, Dec. 2022.
[6] K. Asatani, J. Mori, M. Ochi, and I. Sakata, "Detecting trends in academic research from a citation network using network representation learning," *PLoS ONE*, vol. 13, no. 5, May 2018, Art. no. e0197260.
[7] Y. Zhou, R. Wang, A. Zeng, and Y.-C. Zhang, "Identifying prize-winning scientists by a competition-aware ranking," *J. Informetrics*, vol. 14, no. 3, Aug. 2020, Art. no. 101038.
[8] Z. R. Mahayuddin and A. F. M. Saifuddin Saif, "A comprehensive review towards segmentation and detection of cancer cell and tumor for dynamic 3D reconstruction," *Asia–Pacific J. Inf. Technol. Multimedia*, vol. 9, no. 1, pp. 28–39, Jun. 2020.
[9] S. Zhang, M. Medo, L. Lü, and M. S. Mariani, "The long-term impact of ranking algorithms in growing networks," *Inf. Sci.*, vol. 488, pp. 257–271, Jul. 2019.
[10] J. Zhou, A. Zeng, Y. Fan, and Z. Di, "Ranking scientific publications with similarity-preferential mechanism," *Scientometrics*, vol. 106, no. 2, pp. 805–816, Feb. 2016.
[11] Z. Ma, A. Sun, and G. Cong, "On predicting the popularity of newly emerging hashtags in Twitter," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 64, no. 7, pp. 1399–1410, Jul. 2013.
[12] E. Yan, Y. Ding, and C. R. Sugimoto, "P-Rank: An indicator measuring prestige in heterogeneous scholarly networks," *J. Assoc. Inf. Sci. Technol.*, vol. 62, no. 3, pp. 467–477, Mar. 2011.
[13] Y. Wang, A. Zeng, Y. Fan, and Z. Di, "Ranking scientific publications considering the aging characteristics of citations," *Scientometrics*, vol. 10, pp. 1–12, Jan. 2019.
[14] P. Wouters, C. R. Sugimoto, V. Larivière, M. E. McVeigh, B. Pulverer, S. de Rijcke, and L. Waltman, "Rethinking impact factors: Better ways to judge a journal," *Nature*, vol. 569, no. 7758, pp. 621–623, May 2019.
[15] C. Zhang, "A proposal for calculating weighted citations based on author rank," *EMBO Rep.*, vol. 10, no. 5, pp. 416–417, May 2009.
[16] E. Yan and Y. Ding, "Weighted citation: An indicator of an article's prestige," *J. Assoc. Inf. Sci. Technol.*, vol. 61, no. 8, pp. 1635–1643, 2010.
[17] J. Wilsdon, *The Metric Tide: Independent Review of the Role of Metrics in Research Assessment and Management*. Newbury Park, CA, USA: Sage, 2016.
[18] K. Subramanyam, "Bibliometric studies of research collaboration: A review," *J. Inf. Sci.*, vol. 6, no. 1, pp. 33–38, Jan. 1983.
[19] E. Garfield, "Is citation analysis a legitimate evaluation tool?" *Scientometrics*, vol. 1, no. 4, pp. 359–375, May 1979.
[20] S. Amanlou, M. K. Hasan, and K. A. A. Bakar, "Lightweight and secure authentication scheme for IoT network based on publish–subscribe fog computing model," *Comput. Netw.*, vol. 199, Nov. 2021, Art. no. 108465.
[21] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Netw. ISDN Syst.*, vol. 30, nos. 1–7, pp. 107–117, Apr. 1998.
[22] P. Cui, X. Wang, J. Pei, and W. Zhu, "A survey on network embedding," *IEEE Trans. Knowl. Data Eng.*, vol. 31, no. 5, pp. 833–852, May 2019, doi: 10.1109/TKDE.2018.2849727.
[23] L. Waltman, "A review of the literature on citation impact indicators," *J. Informetrics*, vol. 10, no. 2, pp. 365–391, May 2016.
[24] K. W. Boyack and R. Klavans, "Multiple dimensions of journal specificity: Why journals can't be assigned to disciplines," in *Proc. 13th Conf. Int. Soc. Scientometrics Informetrics*, 2011, pp. 123–133.
[25] L. C. Freeman, "A set of measures of centrality based on betweenness," *Sociometry*, vol. 40, no. 1, pp. 35–41, Mar. 1977. [Online]. Available: http://www.jstor.org/stable/3033543

[26] P. Bonacich, "Power and centrality: A family of measures," *Amer. J. Sociol.*, vol. 92, no. 5, pp. 1170–1182, Mar. 1987.
[27] A.-L. Barabaási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, Oct. 1999.
[28] K. Ko, K. J. Lee, and C. Park, "Rethinking preferential attachment scheme: degree centrality versus closeness centrality," *Connections*, vol. 28, no. 1, pp. 4–15, 2008.
[29] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proc. Nat. Acad. Sci. USA*, vol. 102, no. 46, pp. 16569–16572, Nov. 2005.
[30] S. X. Zhao, R. Rousseau, and F. Y. Ye, "H-degree as a basic measure in weighted networks," *J. Informetrics*, vol. 5, no. 4, pp. 668–677, Oct. 2011.
[31] A. Korn, A. Schubert, and A. Telcs, "Lobby index in networks," *Phys. A, Stat. Mech. Appl.*, vol. 388, no. 11, pp. 2221–2226, Jun. 2009.
[32] A. Abbasi, "H-type hybrid centrality measures for weighted networks," *Scientometrics*, vol. 96, no. 2, pp. 633–640, Aug. 2013.
[33] H. Deng, M. R. Lyu, and I. King, "A generalized co-HITS algorithm and its application to bipartite graphs," in *Proc. 15th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Jun. 2009, pp. 239–248.
[34] M. S. Mariani, M. Medo, and Y.-C. Zhang, "Ranking nodes in growing networks: When PageRank fails," *Sci. Rep.*, vol. 5, no. 1, p. 16181, Nov. 2015.
[35] A. Kumar, S. S. Singh, K. Singh, and B. Biswas, "Link prediction techniques, applications, and performance: A survey," *Phys. A, Stat. Mech. Appl.*, vol. 553, Sep. 2020, Art. no. 124289.
[36] E. Bütün and M. Kaya, "Predicting citation count of scientists as a link prediction problem," *IEEE Trans. Cybern.*, vol. 50, no. 10, pp. 4518–4529, Oct. 2020.
[37] A. C. Tsoi, G. Morini, F. Scarselli, M. Hagenbuchner, and M. Maggini, "Adaptive ranking of web pages," in *Proc. 12th Int. Conf. World Wide Web*, New York, NY, USA, 2003, pp. 356–365.
[38] H. Chang, D. Cohn, and A. McCallum, "Learning to create customized authority lists," in *Proc. 17th Int. Conf. Mach. Learn.* San Francisco, CA, USA: Morgan Kaufmann, 2000, pp. 127–134.
[39] B. Gao, T.-Y. Liu, W. Wei, T. Wang, and H. Li, "Semi-supervised ranking on very large graphs with rich metadata," in *Proc. 17th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, Aug. 2011, pp. 96–104, doi: 10.1145/2020408.2020430.
[40] C.-C. Hsu, Y.-A. Lai, W.-H. Chen, M.-H. Feng, and S.-D. Lin, "Unsupervised ranking using graph structures and node attributes," in *Proc. 10th ACM Int. Conf. Web Search Data Mining*, New York, NY, USA, Feb. 2017, pp. 771–779, doi: 10.1145/3018661.3018668.
[41] M. E. J. Newman, "The first-mover advantage in scientific publication," *Europhys. Lett.*, vol. 86, no. 6, p. 68001, Jun. 2009.
[42] F. Radicchi, S. Fortunato, and C. Castellano, "Universality of citation distributions: Toward an objective measure of scientific impact," *Proc. Nat. Acad. Sci. USA*, vol. 105, no. 45, pp. 17268–17272, Nov. 2008.
[43] F. Radicchi and C. Castellano, "Rescaling citations of publications in physics," *Phys. Rev. E, Stat. Phys. Plasmas Fluids Relat. Interdiscip. Top.*, vol. 83, no. 4, Apr. 2011, Art. no. 046116.
[44] M. S. Mariani, M. Medo, and Y.-C. Zhang, "Identification of milestone papers through time-balanced network centrality," *J. Informetrics*, vol. 10, no. 4, pp. 1207–1223, Nov. 2016.
[45] A. Zeng, S. Gualdi, M. Medo, and Y.-C. Zhang, "Trend prediction in temporal bipartite networks: The case of Movielens, Netflix, and Digg," *Adv. Complex Syst.*, vol. 16, Aug. 2013, Art. no. 1350024.
[46] K. Abbas, M. Shang, A. Abbasi, X. Luo, J. J. Xu, and Y.-X. Zhang, "Popularity and novelty dynamics in evolving networks," *Sci. Rep.*, vol. 8, no. 1, pp. 1–12, Apr. 2018.
[47] K. Abbas, M. Shang, X. Luo, and A. Abbasi, "Emerging trends in evolving networks: Recent behaviour dominant and non-dominant model," *Phys. A, Stat. Mech. Appl.*, vol. 484, pp. 506–515, Oct. 2017.
[48] L. Long, K. Abbas, N. Ling, and S. J. Abbas, "Ranking nodes in temporal networks: Eigen value and node degree growth based," in *Proc. 2nd Int. Conf. Image Process. Mach. Vis.*, Aug. 2020, pp. 146–153.
[49] Y. LeCun, Y. Bengio, and G. Hinton, "Deep learning," *Nature*, vol. 521, no. 7553, p. 436, 2015.
[50] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. W. M. van der Laak, B. van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.
[51] S. Dong, P. Wang, and K. Abbas, "A survey on deep learning and its applications," *Comput. Sci. Rev.*, vol. 40, May 2021, Art. no. 100379.
[52] Y. Bengio, Y. LeCun, and G. E. Hinton, "Deep learning for AI," *Commun. ACM*, vol. 64, pp. 58–65, Jun. 2021.
[53] F. Scarselli, M. Gori, A. C. Tsoi, M. Hagenbuchner, and G. Monfardini, "The graph neural network model," *IEEE Trans. Neural Netw.*, vol. 20, no. 1, pp. 61–80, Dec. 2009.
[54] B. Škrlj, J. Kralj, J. Konc, M. Robnik-Šikonja, and N. Lavrač, "Deep node ranking for neuro-symbolic structural node embedding and classification," 2019, *arXiv:1902.03964*.
[55] X. Wang, P. Cui, J. Wang, J. Pei, W. Zhu, and S. Yang, "Community preserving network embedding," in *Proc. 1st AAAI Conf. Artif. Intell.*, 2017, pp. 1–12.
[56] Y. Li, D. Tarlow, M. Brockschmidt, and R. Zemel, "Gated graph sequence neural networks," 2015, *arXiv:1511.05493*.
[57] M. Defferrard, X. Bresson, and P. Vandergheynst, "Convolutional neural networks on graphs with fast localized spectral filtering," in *Proc. Adv. Neural Inf. Process. Syst.*, 2016, pp. 3844–3852.
[58] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-based learning applied to document recognition," *Proc. IEEE*, vol. 86, no. 11, pp. 2278–2324, Jan. 1998.
[59] S. Bhagat, G. Cormode, and S. Muthukrishnan, "Node classification in social networks," in *Social Network Data Analytics*. Cham, Switzerland: Springer, 2011, pp. 115–148.
[60] Y.-A. Lai, C.-C. Hsu, W.-H. Chen, M.-Y. Yeh, and S.-D. Lin, "DeepRank: Improving unsupervised node ranking via link discovery," *Data Mining Knowl. Discovery*, vol. 33, no. 2, pp. 474–498, Mar. 2019.
[61] Y.-A. Lai, C.-C. Hsu, W. H. Chen, M.-Y. Yeh, and S.-D. Lin, "PRUNE: Preserving proximity and global ranking for network embedding," in *Advances in Neural Information Processing Systems*, vol. 30, I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, Eds. Red Hook, NY, USA: Curran Associates, 2017.
[62] J. Bromley, I. Guyon, Y. Lecun, E. Sckinger, and R. Shah, "Signature verification using a Siamese time delay neural network," in *Proc. Adv. Neural Inf. Process. Syst.*, 1993, pp. 1–12.
[63] W. Glänzel and A. Schubert, "A new classification scheme of science fields and subfields designed for scientometric evaluation purposes," *Scientometrics*, vol. 56, no. 3, pp. 357–367, 2003.
[64] C. Chen, "Predictive effects of structural variation on citation counts," *J. Amer. Soc. Inf. Sci. Technol.*, vol. 63, no. 3, pp. 431–449, Mar. 2012.
[65] E. S. Vieira and J. A. N. F. Gomes, "A comparison of Scopus and web of science for a typical university," *Scientometrics*, vol. 81, no. 2, pp. 587–600, Nov. 2009.
[66] I. Tahamtan, A. S. Afshar, and K. Ahamdzadeh, "Factors affecting number of citations: A comprehensive review of the literature," *Scientometrics*, vol. 107, no. 3, pp. 1195–1225, Jun. 2016.
[67] H. S. Bhat, L.-H. Huang, S. Rodriguez, R. Dale, and E. Heit, "Citation prediction using diverse features," in *Proc. IEEE Int. Conf. Data Mining Workshop (ICDMW)*, Nov. 2015, pp. 589–596.
[68] S. Li, W. X. Zhao, E. J. Yin, and J.-R. Wen, "A neural citation count prediction model based on peer review text," in *Proc. Conf. Empirical Methods Natural Lang. Process. 9th Int. Joint Conf. Natural Lang. Process. (EMNLP-IJCNLP)*, 2019, pp. 4914–4924.
[69] G. Wang, Q. Peng, Y. Zhang, and M. Zhang, "What have we learned from OpenReview?" *World Wide Web*, vol. 26, no. 2, pp. 683–708, Mar. 2023, doi: 10.1007/s11280-022-01109-z.
[70] C.-T. Li, Y.-J. Lin, R. Yan, and M.-Y. Yeh, "Trend-based citation count prediction for research articles," in *Advances in Knowledge Discovery and Data Mining*, T. Cao, E.-P. Lim, Z.-H. Zhou, T.-B. Ho, D. Cheung, and H. Motoda, Eds. Cham, Switzerland: Springer, 2015, pp. 659–671.
[71] X. Ruan, Y. Zhu, J. Li, and Y. Cheng, "Predicting the citation counts of individual papers via a BP neural network," *J. Informetrics*, vol. 14, no. 3, Aug. 2020, Art. no. 101039.
[72] D. Fiala, "Time-aware PageRank for bibliographic networks," *J. Informetrics*, vol. 6, no. 3, pp. 370–388, Jul. 2012.
[73] Y. Zhou, A. Zeng, and W.-H. Wang, "Temporal effects in trend prediction: Identifying the most popular nodes in the future," *PLoS ONE*, vol. 10, no. 3, Mar. 2015, Art. no. e0120735.
[74] D. J. D. S. Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, pp. 510–515, Jul. 1965.
[75] Q. Ke, E. Ferrara, F. Radicchi, and A. Flammini, "Defining and identifying sleeping beauties in science," *Proc. Nat. Acad. Sci. USA*, vol. 112, no. 24, pp. 7426–7431, Jun. 2015.
[76] D. Fiala, F. Rousselot, and K. Ježek, "PageRank for bibliographic networks," *Scientometrics*, vol. 76, no. 1, pp. 135–158, Jul. 2008.

[77] M. Awad and R. Khanna, "Support vector regression," in *Efficient Learning Machines*. Cham, Switzerland: Springer, 2015, pp. 67–80.
[78] S. Eger, P. Youssef, and I. Gurevych, "Is it time to swish? Comparing deep learning activation functions across NLP tasks," 2019, *arXiv:1901.02671*.
[79] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proc. 14th Int. Conf. Artif. Intell. Statist.*, 2011, pp. 315–323.
[80] P. Ramachandran, B. Zoph, and Q. V. Le, "Searching for activation functions," in *Proc. 6th Int. Conf. Learn. Represent.*, Vancouver, BC, Canada, Apr. 2018. [Online]. Available: https://openreview.net/forum?id=Hkuq2EkPf
[81] R. K. Pasumarthi, S. Bruch, X. Wang, C. Li, M. Bendersky, M. Najork, J. Pfeifer, N. Golbandi, R. Anil, and S. Wolf, "TF-ranking: Scalable TensorFlow library for learning-to-rank," in *Proc. 25th ACM SIGKDD Int. Conf. Knowl. Discovery Data Mining*, A. Teredesai, V. Kumar, Y. Li, R. Rosales, E. Terzi, and G. Karypis, Eds. Anchorage, AK, USA, Jul. 2019, pp. 2970–2978, doi: 10.1145/3292500.3330677.
[82] A. Khushnood, "Arxiv citation data processed," Tech. Rep., Jan. 2018, doi: 10.17632/csm9vrmhzd.1.
[83] D. P. Kingma and J. Ba, "Adam: A method for stochastic optimization," 2014, *arXiv:1412.6980*.
[84] J. A. Hanley and B. J. McNeil, "The meaning and use of the area under a receiver operating characteristic (ROC) curve," *Radiology*, vol. 143, no. 1, pp. 29–36, 1982.
[85] J. L. Herlocker, J. A. Konstan, L. G. Terveen, and J. T. Riedl, "Evaluating collaborative filtering recommender systems," *ACM Trans. Inf. Syst.*, vol. 22, no. 1, pp. 5–53, Jan. 2004.
[86] K. Järvelin and J. Kekäläinen, "Cumulated gain-based evaluation of IR techniques," *ACM Trans. Inf. Syst.*, vol. 20, no. 4, pp. 422–446, Oct. 2002.

KHUSHNOOD ABBAS received the bachelor's and master's degrees from Aligarh Muslim University, India, and the Ph.D. degree from the School of Computer Science and Technology, University of Electronic Science and Technology, in 2018. Currently, he is an Assistant Professor with the School of Computer Science and Technology, Zhoukou Normal University, Henan, China. His current research interest includes data mining especially social and information networks.

MOHAMMAD KAMRUL HASAN (Senior Member, IEEE) received the Ph.D. degree in electrical and communication engineering from the Faculty of Engineering, International Islamic University, Malaysia, in 2016. He is currently a Senior Lecturer with the Faculty of Information Science and Technology, Center for Cyber Security, Universiti Kebangsaan Malaysia (UKM). He has published more than 150 indexed papers in ranked journals and conference proceedings. His current research interests include elements pertaining to cutting-edge information-centric networks, computer networks, data communication and security, mobile network and privacy protection, cyber-physical systems, the Industrial IoT, transparent AI, and electric vehicle networks. He is a member of the Institution of Engineering and Technology and the Internet Society. He is a Certified Professional Technologist (P.Tech/Ts.), Malaysia. He served as the Chair for the IEEE Student Branch, from 2014 to 2016. He has actively participated in many events/workshops/training for the IEEE and IEEE Humanity programs in Malaysia. He was an Editorial Member in many prestigious high-impact journals, such as IEEE, IET, Elsevier, Frontier, and MDPI, and the general chair, the co-chair, and a speaker for conferences and workshops for the shake of society and academy knowledge building, sharing, and learning. He has been contributing and working as a volunteer for underprivileged people for the welfare of society.

ALIREZA ABBASI is currently the Director of postgraduate studies with the School of Engineering and IT, UNSW Canberra. His current research interests include enhancing the better functioning and performance of complex socio-technical systems, which primarily lies in the emerging field of social computing, and leveraging the capacity to collect and analyze data to reveal patterns of individual and group behavior.

UMI ASMA MOKHTAR is currently a Senior Lecturer of information science with the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia. She is the Malaysian Team's Lead Researcher for the Inter PARES Trust AI Project. Her articles have appeared in international and national periodicals, such as the *International Journal of Information Management* and the *Records Management Journal*. Her current research interests include electronic records management, function-based classification, information policy, and information security.

ASIF KHAN (Member, IEEE) received the B.Sc. (Hons.) and M.C.A. degrees from Aligarh Muslim University, India, and the Ph.D. degree (Hons.) in computer science and technology from the University of Electronic Science and Technology of China (UESTC), China, in 2016. He was an Adjunct Faculty with the University of Bridgeport, USA, for the China Program, in Summer 2016. Previously, he was a Visiting Scholar of big data mining and application with the Chongqing Institute of Green and Intelligent Technology (CIGIT), Chinese Academy of Sciences, Chongqing, China. He was a Postdoctoral Scientific Research Fellow with UESTC. He is currently an Assistant Professor with Integral University, India. He is a contributor to many international journals with robotics and vision analyses about the contemporary world in his articles. His current research interests include machine learning, robotics vision, and new ideas regarding vision-based information critical theoretical research. He was awarded the UESTC Academic Achievement Award and the Excellent Performance Award, from 2015 to 2016.

SITI NORUL HUDA SHEIKH ABDULLAH received the degree in computing from the University of Manchester Institute of Science and Technology, U.K., the master's degree in artificial intelligence from Universiti Kebangsaan Malaysia, and the Ph.D. degree in computer vision from the Faculty of Electrical Engineering, Universiti Teknologi Malaysia. While starting her career, she was involved in conducting national and international activities, such as Cyber Security Academia Malaysia (CSAM), FIRA, Asian Foundation, IMWU, DFRS, MIAMI, MACE, and IDB Alumni. After serving as the Chairperson of the Center for Cyber Security, she is currently the Deputy Dean of research and innovation with the Faculty of Information Science and Technology, Universiti Kebangsaan Malaysia, and the Chairperson of CSAM. She has published four books titled *Pattern Recognition*, *Computational Intelligence in Data Science Application*, *Smart Prediction of Suspect's Serial Crime Location*, and *Optical Script Reader: Texture Binary Innovation*, and more than 50 journals and 100 conference manuscripts correspondingly. Her current research interests include digital forensics, pattern recognition, and computer vision surveillance systems.

SHI DONG (Member, IEEE) received the M.S. degree in computer science from the University of Electronic and Technology of China, in 2009, and the Ph.D. degree in computer science from Southeast University. He was a Postdoctoral Researcher with the Huazhong University of Science and Technology. He was a Visiting Scholar with Washington University, St. Louis. He is currently a Distinguished Professor with Zhoukou Normal University. His current research interests include network management and network security.

SHAYLA ISLAM (Senior Member, IEEE) received the B.Sc. degree in computer science and engineering from International Islamic University Chittagong, Bangladesh, the M.Sc. degree from the Department of Electrical and Computer Engineering (ECE), International Islamic University Malaysia (IIUM), in 2012, and the Ph.D. degree in engineering from the ECE Department, IIUM, in 2016, under the Malaysian International Scholarship (MIS). She is currently an Assistant Professor with UCSI University, Malaysia. She has awarded the Silver Medal for her research work at IIUM. In consequences, she has also awarded the Young Scientist Award for the contribution of research paper at the Second International Conference on Green Computing and Engineering Technologies (ICGCET'16), in 2016, organized by the Department of Energy Technology, Aalborg University, Esbjerg, Denmark.

DABIAH ALBOANEEN received the M.Sc. degree (Hons.) in advanced computer networking and the Ph.D. degree in cloud computing and artificial intelligence from Glasgow Caledonian University, in 2013 and 2019, respectively. Currently, she is an Assistant Professor with the College of Science and Humanities in Jubail, University of Imam Abdulrahman Bin Faisal. She was awarded a lot of prizes, such as the Best Master Project in Networking and Wireless Communication Prize of the Glasgow Caledonian University, in 2013, and 11 distinction awards from Saudi Cultural Bureau in London, since 2013.

FATIMA RAYAN AWAD AHMED received the B.S. degree in computer science from the Sudan University of Science and Technology, in 2004, and the M.S. and Ph.D. degrees in computer science from Al Neelain University, Sudan, in 2007 and 2012, respectively. She joined Sattam University, Saudi Arabia, in 2013, as an Assistant Professor with the Information Systems Department, from 2013 to 2016, and has been with the Computer Science Department, since 2017. In 2004, she joined Sudanese Company for Telecommunications, Sudan, as a Computer Programmer with the IT Department, where she analyzed, designed, and programmed a set of systems. Her current research interests include artificial intelligence, systems and algorithms analysis and design, web applications, and e-learning.