IEEE Access
Multidisciplinary ⋮ Rapid Review ⋮ Open Access Journal

Received 23 May 2025, accepted 25 June 2025, date of publication 1 July 2025, date of current version 10 July 2025.

Digital Object Identifier 10.1109/ACCESS.2025.3584856

RESEARCH ARTICLE

DiffusionScore: A Framework for Assessing Institutional Research Impact Through Influence Diffusion

GAMBHIRE SWATI SAMPATRAO[1], ANISHA KHETAN[1], NITHIN NAGARAJ[2], (Senior Member, IEEE), SNEHANSHU SAHA[3], (Senior Member, IEEE), SUDEEPA ROY DEY[1], AND ARCHANA MATHUR[4], (Senior Member, IEEE)

[1]Department of Computer Science and Engineering, PES University, Bengaluru 560100, India
[2]Complex Systems Program, National Institute of Advanced Studies, Bengaluru 560012, India
[3]APPCAIR and CSIS, BITS Pilani K. K. Birla Goa Campus, Goa 403726, India
[4]Department of Information Science and Engineering, Nitte (Deemed to be University), Nitte Meenakshi Institute of Technology, Bengaluru 560064, India

Corresponding author: Archana Mathur (archana.mathur@nmit.ac.in)

ABSTRACT Assessing an institution's research output has traditionally relied mostly on the opinion of peers, but as the field of scientometrics progressed dramatically in recent years, the evolution of quantitative indicators gained popularity and became essential. The present approaches for evaluating an institution's research output rely on citation data, that has been quantitatively analyzed and are predominantly data-centric models. It is observed that research productivity and impact are taken into account by many ranking systems, and the measurement of research impact is mostly dependent on citation-counting techniques and the volume of research produced. The recent ranking systems evaluate the reputation of academic institutions by employing network-based algorithms, like Page Rank which examine the citation network at the institution level. An Article's outreach and causal impact are not completely taken into consideration by the aforementioned approaches which leads to incomplete quantification of metrics. In this study, we propose a framework that uses a data-agnostic influence diffusion model to measure the academic impact of the institutions in the citation network. The diffusion model is inspired by the phenomenon of heat diffusion. The suggested framework investigates the citation trajectory for each article published by the university in the research area, analyses its influence diffusion in citation networks, and provides methods, metrics, and processes that others could apply to similar problems or datasets. By leveraging transfer entropy-based techniques, our research goes beyond citation mapping to investigate the causal link between academic influence and the reputation of publication venues employed by universities.

INDEX TERMS Scientometrics, institution research impact, influence diffusion, citation networks, graph theory.

I. INTRODUCTION

Recent years have marked a stark difference in the way research productivity is measured for any institution. The advent of highly sophisticated bibliometric indicators has increasingly been used to estimate research performance at all levels, encompassing articles, authors, institutions, and countries. Researchers and bibliometricians have tried to probe into various methods to evaluate the research performance of an institution by focusing on the research output. To design the proposed framework, we have explored the existing techniques of ranking academic research and investigated the various approaches like PageRank algorithms and its variants, centrality measures, and diffusion or propagation techniques.

The associate editor coordinating the review of this manuscript and approving it for publication was Fu Lee Wang.

According to [1] the Institutional Research Performance (IRP) ought to serve as the foundation for evaluating an institution's performance. The number of full-time faculty members and three institutional variables–such as the volume of published books, journals, and book chapters–as well as case studies are the three institutional variables that the authors proposed to quantify the research output of educational institutions.

As shown in [2], the evaluative investigations focus on the Indian vista in an organized manner. It also highlights the researchers' ongoing efforts to improve their work by incorporating more recent concepts that have been developed throughout time. There has been an honest endeavor to map the quantum of knowledge about this new field of scientometric research. Highlighting the gaps and weaknesses in this field of study is symptomatic, though, and so makes the issues–both attended and unattended–clear. As a result, it provides a logical depiction of this confusing issue with scientific measurement. Further as shown in [3], measures such as size-dependency and free-scale networks were introduced to estimate the scientific contribution of institutions and countries. Their work was a qualitative approach with parameters like - Global Research Output (GRO), which was defined as the number of citations received per publication and a Global Research Output research field (GROr.) which was research field-related measure. The authors further argued that the combination of both types of data (quantitative and qualitative) does help decision-making and policy– making bodies of the institutions. Another breakthrough was using Page rank algorithms to measure academic reputations using citation networks. Many others like [4] used citations to compute the reputation by analysis of citation networks. The mining of such networks leads to interesting facts like influential authors and authoritative center points. With the advent of network metrics like centrality measures, many researchers put eigenvectors into use to compute influential nodes in a network. Cao et al. [5] proposed a unique approach called as heterogeneous ranking method using interconnected institution-publication networks. The authors created the institution-publication network and did a random walk. Then for each layer, a random jump probability was calculated. The authors also calculated the inter-collaboration-citation jump to look into the interactions. Finally, extensive experiments were conducted on large-scale empirical data from American Physical Society journals. The results demonstrated that the proposed method, HRank, performs well in identifying influential institutions. Another metric that gained huge popularity was the usage of Eigenvector. Wang et al. [6] proposed a heterogeneous academic network consisting of links between three types of academic entities (authors, papers, and venues). In addition, a new academic influence ranking algorithm, AIRank, is proposed to evaluate papers' academic influence. This approach investigates node pairs based on attributes like citation emotional attribute, semantic similarity, and academic quality differences between node pairs. With the recent advances in the field of bibliometrics indicators many researchers have emphasized the need to redefined research impact. The contribution of research output to facilitate scientific and technical advancement is defined as ''scientific or scholarly impact'' [7]. Along with scholarly impact, measuring Scholastic independence is addressed by [8]. In this paper, authors have proposed a methodology that measures 'scientific independence' by removing local citations to define influence. The metric ALIS was created by identifying all citations from local or known acquaintances using a tree structure called as the genealogy tree. The work was further backed up by graph theoretic approaches and resulted in significant insights related to author citation networks. Recently, there has been a lot of curiosity in discovering nodes with significant influence and ranking them in social networks. Graph centrality metrics such as Betweenness, Degree, Eigenvector, Closeness centrality, and Page Rank are frequently used to identify influential nodes. The simplest technique, known as degree centrality, is determined by the number of linkages that occur on the edges. According to [9] the authors have created a heterogeneous structural measure method to unveil the impact of institution and paper, reflecting the effects of citation, institution, and structural measure. The model's performance is evaluated first by constructing a heterogeneous institution-citation network based on the American Physical Society (APS) dataset. Next PageRank is used to quantify the impact of institutions and papers. Finally, same impact institutions are merged, and the ranking of institutions and papers is calculated. The suggested algorithm by [10] indicates spreading capabilities through neighbor variation and an expanding sphere. The total number of shells in which a node has neighbors is known as its neighborhood diversity. The diversity strength centrality, which takes into consideration the intensity and diversity of the neighbors, is defined. In order to more precisely determine the most critical nodes, the authors of [11] investigated betweenness centrality and closeness centrality based on the overall structure. Further, to evaluate the performance, the authors used a stochastic Susceptible-Infected-Recovered (SIR) information diffusion model to compute various metrics including the infection scale, the final infected scale over time, and the average distance between spreaders. As suggested by Berahmand, Kamal, et al. [12] an efficient approach to improve local random walks in complex networks like the citation networks. The authors used mutual information to encourage random walks to shift toward nodes with greater influence. The influence of the source node was used to decide which node would be chosen next. The approach was found to show improved prediction accuracy in 11 real-world networks when compared to other similarity-based approaches. Scarabaggio et al. [13] proposed a novel approach to select the initial node groups to maximize influence spread in large-scale networks, which intends to remove any overlap of seed node influence spread. Iterative numerical experiments showed that this algorithm finds near-optimal solutions and outperforms state-of-the-art algorithms in large-scale scenarios. Piao et al. [14] approached measuring the influentiality of a node which is based on the user's activity, login details, and blog response as essential influence. The authors claimed that when PageRank is used to determine user impact, there is a considerable lack noticed in each node's individual values. as page rank follows the equal value transfer of followers' influence. This paper uses the propagation network attributes of Weibo, a Chinese service that is similar to Twitter. It then empirically quantifies the contribution value of followers' influence to the users they follow centered on various interaction influences. Furthermore, we evaluate how relevant users' individual interest choices and topic content are, and we enable real-time tracking of users' influence at multiple points in time during the public opinion dissemination process. Dey et al. [15] applied graph centrality measures and page rank to the co-authorship Network and used epidemic models to compute the diffusion score. The use of epidemiology has been a profound breakthrough as diffusion scores are computed based on the various states at which the author or institution is. The Susceptible, infected, and recovered stages are suitably modified to encompass the parameters to justify influentiality. In this study [16] the use of network centrality measures was further leveraged to identify impactful articles within a journal. Authors further investigate the citation count, Pagerank, and a local diffusion method to act as an effective filter and correlate well with citation count within a journal. Their study further shows the flow of citations in the article citation network. The article further investigates the information flow that occurs between venue-generated communities. The inferences drawn showed that to measure the true impact of any article one should consider both direct and indirect citations. Visualizing such networks has been an important aspect as key indicators can be suitably found. Further in [17] it was discussed that data and techniques existing in bibliometrics can be used either by itself or in collaboration with qualitative approaches to analyze the knowledge flows. This work categorizes the study of knowledge flows based on two kinds of flow: interpersonal or impersonal. The authors further categorize the object of research as formal, informal, or tacit knowledge. The paper points out highlights that the existing bibliometric approaches have underutilized potential and diffusion methods can be further explored in the realm of real-world scenarios. This work introduced a new metric for ranking academic institutions [18]. This was based on the computation of influence on the global research community. The metric called 'Non-local Research Influence Score' (NLRIS) was calculated using the Constant Elasticity of Substitution (CES) function. The metric performed reasonably well to smaller institutions that perform well in research quality and international collaboration. The extraction of knowledge flows and identifying crucial patterns was explored in [19]. Some important parameters such as knowledge diffusion capacity, knowledge flow intensity, and transfer capacity were employed to compute the prestige and impact of an institution. The citation impact of articles is calculated by using the entropy-weight method to weigh these three patterns.

In this work, we propose a novel bottom-up approach to capture, analyze, and measure the influence spread of a research institution. The lowest level comprises the article and based on the article scores, which is further used to compute the institution-level metric for influence score in a particular domain. The citation network acts as a backbone for the study as it is used to analyze the spread of influence in an article's citation network. Also, for a domain, the article's multiple-hop citation network is constructed, that records the direct and indirect citations received by the article. The investigation is further supplemented with an assumption that publishing a work in a high-SNIP or high h-index journal (or conference) impacts the influence spread of the article and of the affiliating institution.

The rest of the paper is structured as follows. The subsequent section presents the problem definition, our contribution, and the motivation. The "Methodology" section covers the model's mathematical formulation based on the idea of heat diffusion, as well as the preliminary steps, data curation, and creating citation networks. The experiment's setup, data set, and analysis of the results are covered in the section under "Experimental Setup and Result Analysis." We examine the cause-and-effect relationship in the "Causality Analysis" section to identify the elements that increase the impact and dissemination of articles released by a specific organization.

A. MOTIVATION, PROBLEM STATEMENT, AND CONTRIBUTION

Traditionally, the research impact and academic reputation of institutions are computed using citation-count-based metrics such as the h-index, g-index, and R-index [20], [21], [22]. While these citation cased metrics capture the direct citations of research articles, they tend to overlook the underlying structure of the citation network, especially the indirect citations and the diffusion of influence through these citation chains [23], [24]. The limitation of current methods is that they fail to consider the topological structure of citation networks, which can offer deeper insights into the outreach and influence of academic work that in turn can influence diffusion of institutions in a particular research field. In the field of bibliometrics, another approach is very prevalent and lot of researcher's have tried to analyze citation network based on its topological structure. Various ranking algorithms are proposed to evaluate the impact of the article in the citation network [25], [26], [27]. Our work is based on the latter approach and it analyzes the influence spread of an article by considering the topological structure of its citation network.

The intent of this work lies in the fact that there is no framework to comprehensively assess the impact of institutional research. Furthermore, while several citation-based indicators have been explored so far, the use of heat diffusion or heat kernel-based models remains largely unexplored in citation network analysis. This gap motivates us to develop a reachability-based diffusion model that considers the full citation chain of an article. Additionally, we seek to understand the reasons behind the differing performance of institutions by exploring cause-effect relationships through causal analysis, specifically using Transfer Entropy.

Our contribution aims to achieve the following:

1) To develop a data-agnostic framework that quantifies research impact of any institution by suitably capturing the dynamic process of influence diffusion in citation networks, going beyond traditional citation-count analysis and centrality measures. By doing so, it will enable a deeper and fair understanding of an institution's role and impact within the academic ecosystem.
2) We introduce a reachability-based diffusion model that leverages the topological structure of article-level citation networks to quantify influence diffusion. This model accounts for both direct and indirect citations, thus capturing the complete spread of influence from an article.
3) We further extend our analysis by using causal analysis using Transfer Entropy to investigate the cause-effect relationship of research performance across institutions, thereby getting insights into why certain institutions outperform others in research impact.

II. METHODOLOGY

A. PRELIMINARIES

1) CITATION NETWORK REPRESENTED BY A DIRECTED GRAPH

Assume G to be a directed graph given as $G = (V, E)$ where the vertex set V, indicate articles, is given by $V = \{v_1, v_2, \dots v_n\}$, and edges represent citations, from any v_i to v_j } is shown as $E = \{(v_i, v_j)$. An edge from v_i to v_j shows article v_j influences article v_i implying the direction of influence flow is opposite to the direction of citation. Figure 1 represents a citation network of node 0 (source article). This citation graph has 6 nodes (articles). Here, articles 1, 3 and 4 cite article 0 so arrows of influence are opposite. Article 5 cites article 2 which in turn cites article 4. Note the direction of the arrow is opposite to show the influence diffusion. Further the network is used to investigate direct and indirect citations. Since article 0 and article 3 are connected, it indicates article 0 is cited by article 3. We use Article 0 as the starting point for calculating the influence as it has an influence that flows in the direction of the edges. Articles that have cited source articles 0 indicate direct citations (for example, citations from articles 1, 3, and 4), whereas articles have cited articles 1, 3, and 4 are considered indirect citations

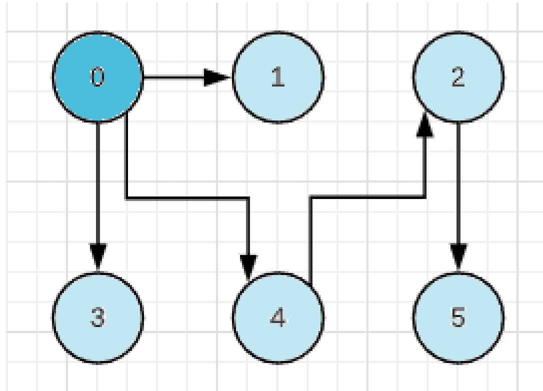

FIGURE 1. Graph (Citation network of the article) T shows influence flow.

for source 0. The idea is to capture and quantify the influence of a source node given any citation network

- The nodes that cite the source node directly are regarded as being in the first layer if we consider the network as a tree. The nodes in the second layer are directly connected to those in the previous layer, and so on. The source node should have a considerable impact on the nodes in the layers that are closest to the source node. Considering that these nodes represent papers that have a direct citation to the source. Eg. in figure 1, since nodes 1, 3, and 4 are located in layer 1, their influence should be significant.
- A node's count of incoming and outgoing edges must also be taken into account for quantifying the amount of influence it has on the nodes. This shows the transfer of influence from a region with more influence to one with less influence. In figure 1, the impact of source node 0 is diffused to its nearby nodes, who then pass the influence on to their neighbors. The degree of the node determines how much influence is kept in the node and how much is disseminated. Intuitively, since it has cited the source node over several direct and indirect paths, the node is more influenced by the source node the more incoming edges it has. The more edges it has on the outside, the more heat it can transmit to its offspring nodes.

The influence diffusion model must be reachability-based as opposed to activation-based models, where a node's effect is determined by the number of inactive nodes it may activate. The relevance of a node is measured by the influence it can diffuse to its neighboring nodes which may reach out through multiple paths. This inspired us to look into whether the concept of the heat diffusion method was feasible to model this phenomenon. We model the influence diffusion of a citation network as a process of heat diffusion. The diffusion of heat is a physical phenomenon. Heat flows from a high-temperature position to a low-temperature in a medium. The partial differential equation can be used to depict the heat flow in a geometric manifold. If heat p is a function of distance x and time t, then $p(x, t)$ can be given as

$$\frac{\partial u}{\partial t} = k \frac{\partial^2 u}{\partial x^2}, \tag{1}$$

A heat kernel, also known as a diffusion kernel, is a solution to the heat equation that depicts heat transmission or diffusion on a graph as a rough approximation of the underlying geometric manifold. The temperature distribution does not change after an equilibrium time, and the system enters a steady state. Heat diffusion is a type of topological diffusion in which diffusion is manifested as topological connections between nodes. In social network analysis, heat diffusion has been applied. In the article [28] authors suggested an activity-based social influence model based on the heat diffusion principle to evaluate influence diffusion in networks. Heat is distributed or received as a result of interaction between nodes. In social networks, the authors looked at both interactive and non-interactive behaviors. Influence coverage is used to calculate the top-k nodes. The article [29] proposes several topological diffusion models to identify the most important nodes, one of which is the heat diffusion Kernel. The author presented a heat diffusion kernel based on non-interacting activity, a heat diffusion kernel based on interactive and non-interactive activities, along with the proposed one. The distance between the nodes and the source node is not taken into account in these articles, hence the number of hops from the source node has no bearing on the quantity of heat transmitted.

The article is represented by the nodes in the citation network, and if there is a link between two nodes, it indicates a citation relationship between them. When an article cites another article (let's say, 'A'), 'A' impacts the citing article, which then allows 'A' to spread its influence. A's influence grows in the citation network as the citing article receives citations, but the amount of influence spread decreases. A heat source in a citation network is a source article, and the dissemination of the influence can be described as a heat diffusion process. We adjusted the heat kernel to reflect the fact that the closer a node is to the source node, the more it is influenced by it. This is because influence dissemination will be greater if the article directly cites the source article. Furthermore, if the node occurs in many hops from the source node, showing that the article is citing the source article via multiple channels, it will earn additional influence. The *diffusibility* of each node in the citation network is computed and used to create the heat kernel.

B. BUILDING CITATION NETWORKS

To build a citation network, we have used AMiner [*DBLP-Citation-network V11* (available at www.aminer.org/citation)], which is a well-known free repository that has been profoundly used to search and extract large scientific data. Further, the metadata contained papers that are in JavaScript object notation format (JSON). After preprocessing, 27 fields such as ID, Title, Author names, year, Keywords, Field of study, References, and Doi were extracted from the data schema. The venue, doctype, and publisher data are also retained for better clarity. To build a citation network, we randomly chose 20 Universities, based on the QS World University Rankings from the domain of Computer Science and Information Systems. These universities ranked between 1 to 550 in the QS world ranking ensuring consistency in the samples taken. Further, 910 articles published in the domain of *Machine Learning* were extracted. We further break down the citation network for each article as follows:

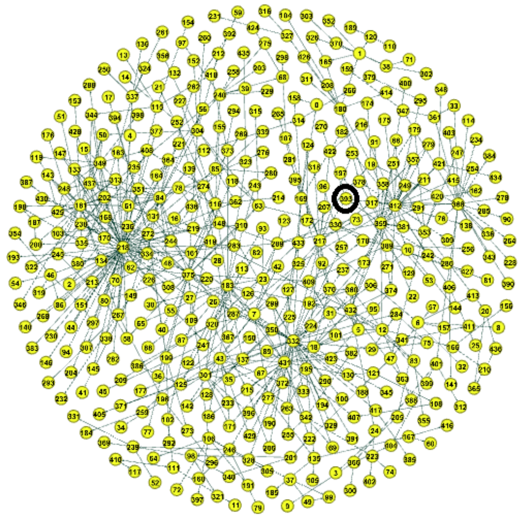

FIGURE 2. Citation network visualization of 440 articles Using Gephi - The source article (in black circle) belongs to UN4.

1) SOURCE AND TARGET PAIR CREATION

Creation of source and target articles was carried out by labeling the source as the base article and the target as those that cited the base. The pair of source and target ids are constructed by collecting the indexes of cited and citing articles, and a citation network of base articles is built by incorporating the number of citation hops/levels into account.

2) ADJACENCY MATRICES CREATION

To ease the implementation and analysis process, the Citation network obtained above is represented in the form of the *adjacency matrix* data structure. Figure 2 represents the radial view of the final citation network for an article. Please refer to Appendix Algorithms 2 and 3 for a detailed explanation of steps taken to create adjacency matrices and citation networks.

3) CAPTURING STRUCTURAL DIVERSITY

We believe graphs can differ widely not just in how big they are (number of nodes/edges), but how the information flows – i.e., the shape or structure of the graph. Our main intention was that if the framework performs well across diverse structural topologies, it will generalize well, even if applied to graphs in new domains, institutions, or regions. This approach is important in the field of scientometrics because most existing models are caught in the trap of capturing data diversity and completely neglecting the importance of structural constructs. To empirically demonstrate the structural diversity inherent in our dataset, we included multiple topological features across the 910 citation networks. This analysis is presented through various tables shown below that capture variations in depth, breadth, and overall network size, offering a multi-dimensional view of structural heterogeneity. Table 1 shows the distribution of citation depth (measured in hops) across networks, ranging from shallow (1-15 hops) to deep structures (>45 hops).

TABLE 1. Structural diversity captured through depth/hops.

Hops in Networks	1 to 15	16 to 30	31 to 45	>45
No. of Articles	720	112	36	42

TABLE 2. Structural diversity captured through breadth of first hop/direct citations.

Direct Citations	1 to 10	11 to 20	21 to 30	>31
No. of Articles	676	125	38	71

Table 2 reports the breadth of the networks as measured by direct citations at the first hop. The spread from 1-10 citations to highly connected (>31 citations) indicates heterogeneity captured in the network structure.

TABLE 3. Structural diversity captured in terms of size through total citing nodes.

Total no. of citing nodes	1-200	201 to 400	401 to 600	>600
No. of Articles	619	69	35	187

Table 3 captures the structural diversity in terms of the size of the network. The various ranges indicate the underlying citation network structures across the 910 articles, demonstrating significant topological diversity.

Although AMiner and DBLP provide comprehensive and carefully curated metadata, we recognize that datasets are not without limitations concerning disciplinary coverage, regional representation, and non-English publications. Some known biases could be overrepresentation of computer science and related fields in DBLP, the underreporting of publications in regional or non-English venues, and potential inconsistencies in author disambiguation.

We make a note here regarding our model, which is data-agnostic.The proposed model is validated on a carefully chosen small set of real article data. It is also nontrivial to mention that the model should be robust enough so that changes in data do not affect the model behavior drastically. We ensure that the model is not fitted from data. Rather, it is a qualitative interpretation of how the scoring mechanism should work. Consequently, it is crucial to understand the model and the computational approximation to determine the outcome from it.

C. THE MODEL

The following are the steps used in this work:

- From the Amner DBLP citation network v11 dataset, the publications and citation data for selected universities are curated once the data has been preprocessed and cleaned as described in the previous sections.
- An adjacency matrix was used to depict the citation networks for the articles in the graph.
- We examine the graph topology and calculate parameters for the Diffusion Model to use.
- We generate a Heat Kernel for the article citation network based on these parameters, which is utilized for computing the influence scores of the articles.
- The institution's influence score is calculated by combining the influence scores of all published articles.

The proposed influence diffusion model is explained in this section. Using the proposed heat kernel, we examine the article's citation network and quantify the article's influence. The article's influence score is the sum of influence spread on all articles in a network. The source node has a positive amount of heat, to begin with. Heat is not present in any of the other nodes. This initial heat from the source will diffuse throughout the network, heating the other nodes. The Diffusability hd for each node is defined in terms of parameters α and β as follows:

The amount of influence a node in the network can disperse to its neighboring nodes is known as its Diffusability. To compute α, we first compute each node's shortest distance from other nodes in the network, then add them up for each node as follows:

$$\alpha[i] = \alpha[i] + 1, \text{ if } R[i][j] > 0, \tag{2}$$

R is the reachability matrix.

$$\alpha[i] = \begin{cases} 1, & \text{if } \alpha[i] = 0, \\ \alpha[i], & \text{otherwise.} \end{cases} \tag{3}$$

The sum of all node distances from $i \in V$ is $\alpha[i]$, indicating the amount of influence a node can diffuse. If a node has a large number of nodes connected to it directly or indirectly, it will diffuse more effectively, resulting in a higher α value for these nodes. The diffusibility of a node in terms of in-degree and out-degree is indicated by the parameter β. To calculate β, multiply the number of edges supplying influence to the node by the number of edges diffusing the influence to the other nodes. As a result, we add the node's entering edges and remove the node's exiting edges as follows:

$$\beta[y] = \begin{cases} 0, & \text{if } d_{in} = 1 \text{ and } d_{out} = 0, \\ d_{in}[y] - d_{out}[y], & \text{otherwise.} \end{cases} \tag{4}$$

Now we define the diffusibility hd of each node using α and β as:

$$hd[i] = hd[v] - \beta[i]/n + \alpha[i]/n. \tag{5}$$

We traverse the graph using the depth first search (DFS) method, calculating *hd* for each node one by one. When the number of incoming edges at a node exceeds the number of outgoing edges, $\beta > 0$, and since the diffusibility of this node should be lower, we subtract β from $hd[i][j]$. If the incoming edges at a node are less than the outgoing edges, the value of β is 0, and since the diffusibility of this node should be higher, we remove β from $hd[i][j]$, which is added to the diffusibility.

The higher α value for a node indicates that the node is connected to many other nodes, directly or indirectly. So, α is added to $hd[i][j]$.

If node j is closer to node i, the value of $hd[i][j]$ is lower. This explains why the influence is greater in nodes closer to the source node and decreases as we travel away.

Now, the influence spread by the source node in the graph can be calculated after determining the diffusibility constant *hd* for each node.

Influence in a node i is calculated as the total of the influence it received from its parent nodes k less it dispersed to all of its direct and indirect links defined as:

$$h(T + \triangle T) - h(T) = \sum_{(v_i, v_j \epsilon V)} (h_k(T) - h_j(T)), \tag{6}$$

where $h_k(T)$ is the influence received from parent nodes k and $h_j(T)$ is the influence diffused from a node i $h_k(T)$ is proportional to $\triangle T$, and inversely proportional to *hd* Since the lesser the value of *hd* more heat will be retained in the node. Additionally $h_k(T)$ is proportional to the constant z_i that is set to 1 if there is an edge between i and k, and otherwise it is 0. Hence,

$$h_j(T) = h(T) \cdot \triangle T \cdot hd[j][i]/n. \tag{7}$$

Equation 6 can be written as:

$$h(T + \triangle T) - h(T) = h(T) \cdot \triangle T \sum_{(v_i, v_j \epsilon V)} \left(\frac{z_i}{hd[k][i]} - \frac{hd[j][i]}{n} \right), \tag{8}$$

When $\triangle \mathrm{T} \rightarrow 0$,

$$\frac{dh(T)}{dl} = \alpha HKh(T), \tag{9}$$

where α is diffusion coefficient. We set $\alpha = 1$. HK is the heat kernel. Solving this equation by taking the limits, we get:

$$h(t) = e^{HK`T} h(0). \tag{10}$$

Equation 11 gives quantified influence at the nodes. The quantity of influence received and dispersed by a node is specified by the heat kernel HK' in Equation 11.

$$HK`[x, y] = \begin{cases} 1/hd[y, x], & \text{if } t[y, x] = 1 \text{ and } hd[y, x] \neq 0, \\ -hd[y, x]/n, & \text{otherwise.} \end{cases} \tag{11}$$

Algorithm 1 Algorithm for Quantifying Influence Diffusion in the Network

Input: Initial heat in a network: f(0)
Output: The source article's Influence Score
Step 1: Define the adjacency matrix and the Parent Matrix
Step 2: Calculate α and β for every node in the network as indicated in equations 3 and 4
Step 3: Combine α and β to calculate *hd*
Step 4: Define *HK*(Heat kernel)
for *each row in graph g* **do**
  for *each column in graph g* **do**
    if *there is an edge between column node and row and value of L for row and column !=0* **then**
      $HK[x, y] = 1/hd[y][x]$
    end if
    else
      $HK[x, y] = -hd[y][x]/N$
    end if
  end for
end for
Step 5: Calculate influence as shown in equation 11.

Let, A be the set of all the articles published by institution I. Let $a_i \in A$ be an article of institution I.

$$Inflscore\_a_i = \sum \frac{\text{Amount of influence present in the layer L}}{\text{Number of nodes in L}} \tag{12}$$

$$Inflscore\_U = \frac{\sum_{a \epsilon A} \text{Inflscore\_ai}}{\text{Average\_source\_heat\_of\_all \_citation\_networks}} \tag{13}$$

When the impact score of all the institution's published articles is added together, Equation 13 gives the institution's influence diffusion score. The algorithm 1 summarizes steps for quantifying influence diffusion in the network. The steps to compute α and β are available in Appendix A (algorithm 4 and algorithm 5)

III. EXPERIMENTAL SETUP AND RESULT ANALYSIS

In this section, an analysis of our model is presented on a single-article citation network. We put our structure to the test in a real-world situation. We extracted data from 910 articles published by 18 universities. The publication year was narrowed down to 2012-2013. Furthermore, citations received through May 2019 for each papers were examined to develop citation networks. We have 910 citation networks, one for each article. We want to measure each article's influence dissemination concerning its citation network.

After assigning some amount of initial heat to the source node, the proposed diffusion model calculates the heat dispersion in the network. The initial heat is the H-index

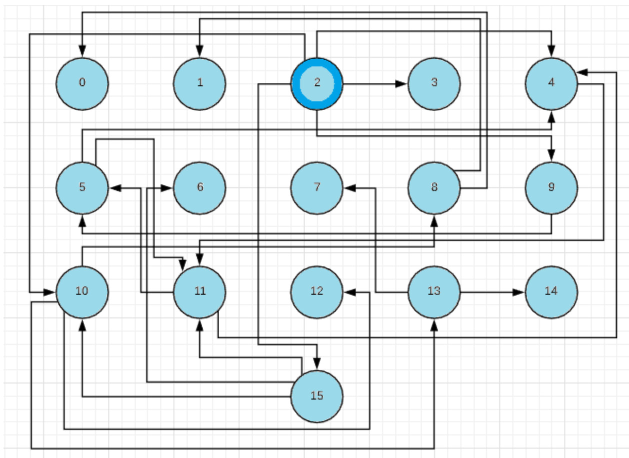

FIGURE 3. Citation network of article id 2050230532 (Node 2) that belongs to UN4.

TABLE 4. Analysis of the network structure and influence diffusion of the article id 2050230532 (Node 2) that belongs to university UN4.

Node Id	Hop No	Quantified Influence Value at Node	In Degree	Out Degree
0	3,4	0.0381	1	0
1	3,4	0.0381	1	0
2	0	0.0415	0	5
3	1	0.0758	1	0
4	1,3,4	0.0758	3	1
5	2,3,4,5,6	0.0387	2	2
6	2	0.0379	1	0
7	3,4	0.0381	1	0
8	2,3	0.0381	1	2
9	1	0.0676	1	1
10	1,2	0.0598	2	3
11	2,3,4,5	0.0388	3	2
12	2,3	0.0383	1	0
13	2,3	0.0381	1	2
14	3,4	0.0381	1	0
15	1	0.0548	1	3

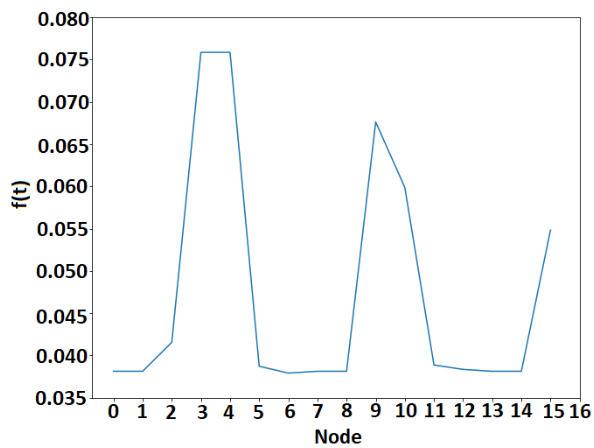

FIGURE 4. Heat/influence present at the nodes in the network.

or SNIP (Source Normalised Impact Per Publication) of the conference or journal where the article is published Normalized values of the H-index and SNIP are provided for the suggested influence diffusion model because their scales are different. The H-index is an author-level metric. It is defined as - 'h' number of papers receiving 'h' citations. At the conference level also, H-index is a well-accepted parameter of quality. To accommodate for discrepancies in citation counts between scientific domains, SNIP uses a normalized citation count. It takes into account the number of publications in the previous three years in addition to the citation count. Heat is initially applied to all other nodes in the citation network with a value of 0.

Figure 3 depicts a network of an article belonging to university with ID UN4 and article id 2050230532 (the source article is designated as 2 in the network). 'International Conference on Mobile Systems, Applications, and Services (ACM)' published this article in 2013. Table 4 depicts the network structure as well as the heat generated at each node.

Source node 2 in the network gets initial heat 0.07584. This is normalized H-index for the source article. Other nodes are initialized with zero initial heat. In figure 3, nodes reachable from source node2 are:

[3, 4, 9, 10, 15] 1 hop
[5, 6, 8, 10, 11, 12, 13] 2 hops
[0, 1, 4, 5, 7, 8, 11, 12, 13, 14] 3 hops
[0, 1, 4, 5, 7, 11, 14] 4 hops
[5, 11] 5 hops
[5] 6 hops

The distribution of heat/influence in the network is determined by the distance between nodes and the source node and the degree of each node. 3, 4, 9, 10, and 15 are the nodes that can be reached on the first hop. Because it is directly citing the source article and is not connected to any other nodes to transfer the heat further, node 3 has the greatest influence. Node 4 has a high influence because it cites the source article several times and has an out-degree, distributing influence to other nodes as well.

The heat/influence diffused in the network node-by-node and layer-by-layer is depicted in Figures 4 and 5. The outcome is consistent with the method used to create the influence diffusion model. The article that directly cites the source article will be heavily influenced, whereas publications that cite indirectly (after the first hop) will be less influenced. In addition, if the article is present in multiple citing paths to the source article, it is influenced heavily as a result. Furthermore, the diffused influence is affected by the article node's indegree and outdegree.

Figure 5 illustrates that layers closer to the source nodes have a higher average influence on the nodes than layers farther away when the number of nodes in the layer is taken into account. Next, we present an analysis of the entire institutional data that contains 910 article citation networks for 18 universities. The presence of linear relationship between the diffusion score and the number of hops was confirmed through scatter plots.

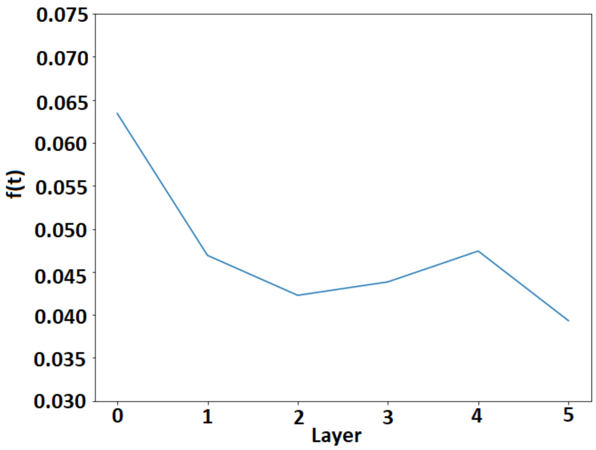

FIGURE 5. Heat/influence present in every layer of the network.

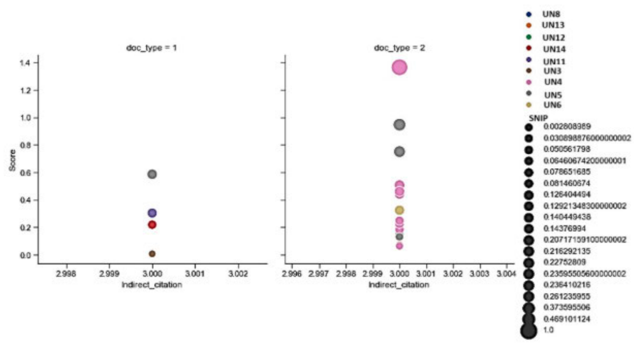

FIGURE 6. Effect of direct citations on scores (with same or comparable indirect citations).

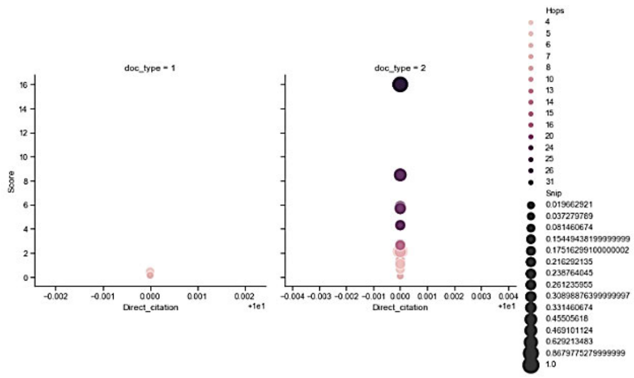

FIGURE 7. Effect of indirect citations on scores (with same or comparable Direct citations).

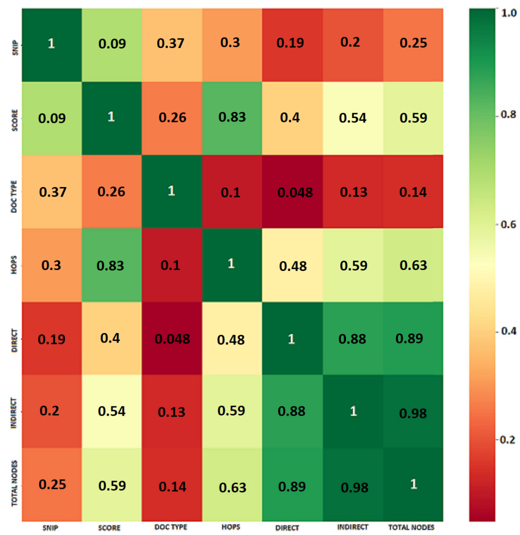

FIGURE 8. Pair plot indicating the strength of a linear relationship: Correlation.

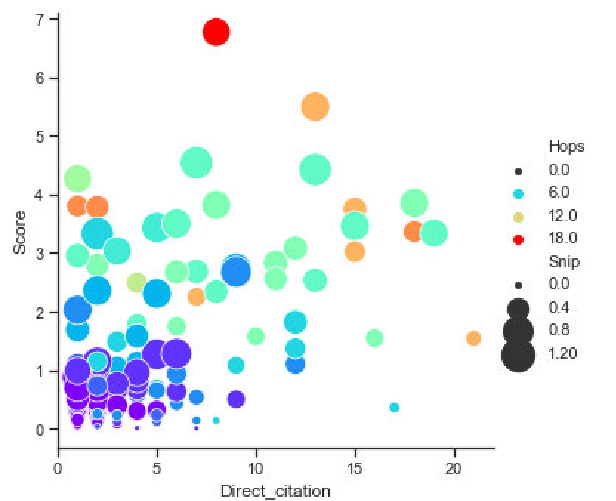

FIGURE 9. Comparison of direct citations vs score with hops and SNIP.

A heat map (Figure 8) showing the level of correlation between the diffusion and all the attributes is created. This was further validated by the pair plot which indicated a significant Pearson Correlation of 0.83 between the two. A significant linear relationship is shown between the diffusion score and the number of hops in the network. To examine the effects of direct (citations obtained by the original article) and indirect citations, we analyze additional attributes and diffusion scores of the universities with similar or comparable citation counts as shown in Figure 6 and Figure 7. Comparison results demonstrate, diffusion scores are not solely dependent on raw citations but tend to rise with higher hops and SNIP/H-index values.

Figure 9 illustrates the influence of direct citation count on the diffusion score using a bubble chart with various hues for SNIP/H-index and hops. Because the points are dispersed randomly, there is little to no association between diffusion score and direct citation. The scores rise as both the number of hops and the snip value do. It means that the influence diffusion score isn't exclusively determined by the number of citations it receives (direct citations). Next, it is looked at how the diffusion score of the source article is affected by indirect citations, as seen in figure 10. These are the multiple-hop citations to the source article. The diffusion score rises as the number of indirect citations received by the

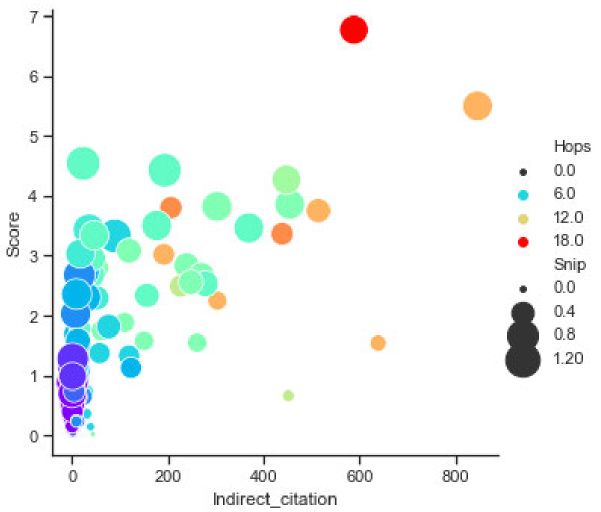

FIGURE 10. Comparison of indirect citations vs score with hops and SNIP.

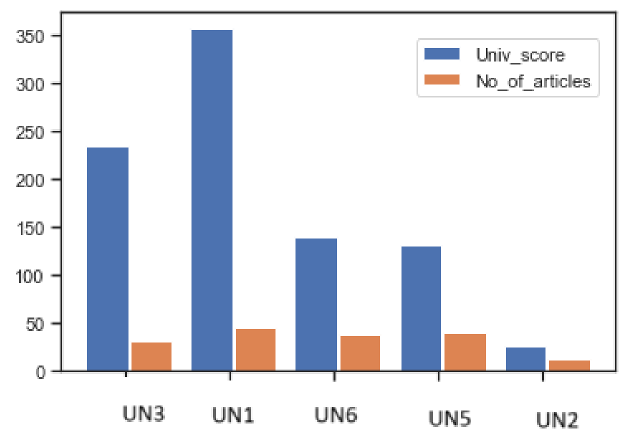

FIGURE 11. University score comparison.

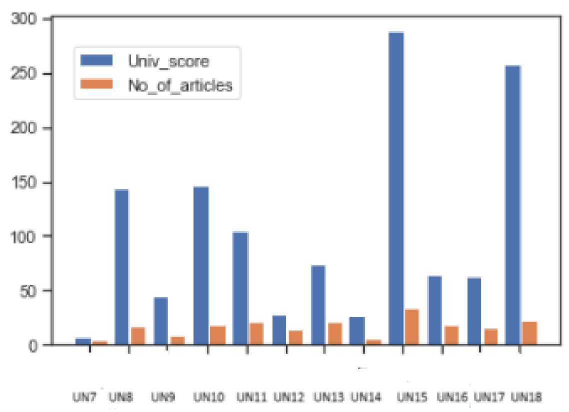

FIGURE 12. University score comparison.

source article rises, as does the SNIP/H-index and the number of hops. We conclude that, rather than just the raw number of citations, the article's influence diffusion is influenced by its citation network structure, depth, and the journal's or conference's SNIP/H-index. The *Influence Diffusion Score* for the 18 universities under study is shown in Table 5 and Figure 12. In the case of two colleges, UN13 and UN11, both of which have the same volume of publications (21), they have gotten 152 and 105 citations, respectively. UN11, has a higher influence diffusion score than UN13. This is because the proposed model quantifies the diffusion of influence in the citation network, which is dependent on the publishing body's network topology and SNIP/H-index. The citation networks of UN11's publications have a higher number of indirect citations (11164 total), showing that its influence is more widely dispersed since the breadth and depth of citations obtained is greater, whereas UN13 has 2101 indirect citations. Furthermore, UN11 average SNIP/H-index is 0.2032, whereas UN13 is 0.1668. Few more runs were also conducted to ensure our dataset spans across a broad spectrum of venues' prestige as well, by using popular metrics like H-index and SNIP. This further reinforces that the observed diversity in network topologies is generalizable across heterogeneous academic ecosystems. The results of these runs are added to the section Appendix as Tables 6, 7, 8 and 9.

A. CAUSALITY ANALYSIS

The quantification of research influence is a significant issue that has been studied from a variety of angles. However, there is no casual examination of influence in the literature. Unsubstantiated claims abound, such as the idea that publishing in reputable scientific journals can assist individuals and institutions gain more influence. The following hypothesis is being investigated: The scholarly impact of articles is enhanced if it is published in a conference

TABLE 5. Influence diffusion scores for 18 universities.

University	Articles	Direct citations	Diffusion-Score
UN1	83	610	700.2923
UN2	24	174	79.1337
UN3	91	970	534.0466
UN4	322	3727	2422.5427
UN5	107	815	726.4202
UN6	81	731	641.0037
UN7	4	8	7.0407
UN8	17	78	144.4719
UN9	9	57	45.4662
UN10	19	139	146.9682
UN11	21	105	104.8084
UN12	14	43	27.9011
UN13	21	152	74.2840
UN14	6	30	26.4534
UN15	34	352	288.6475
UN16	18	98	64.7140
UN17	16	150	62.7385
UN18	23	347	258.4464

or journal with a sufficiently high h-index. Institutions that publish in such venues regularly are naturally positioned in higher ranks of scholarly prominence. We investigate the

Algorithm 2 Creating Pairs of Base Articles and Citing Articles

Input: Preprocessed CSV file
Output: Pairs of base article and citing articles
1: Load CSV.
2: Select the required columns 'univ','publisher','year','title','fos','references','doctype','id'.
3: Convert the values of string type to list.
4: Create *CitedBy* and *Ind_CitedBy* as new columns. **for** *row(i) in a file* **do**
  Build *index_l* and *ctdlist*.
  for *each row r(index) in file* **do**
    if *i[id] in index[references]* **then**
      Add index to *index_l*.
      Add index[id] to *ctdlist*.
    end if
    Add the *index_l* to *Ind_CitedBy* column.
    Add *ctdlist* to *CitedBy* column.
  end for
end for
5: Construct a Network of indices
Run Algorithm 3.
6: Construct Source and Target pairs where, Source: Base article, Target: Citing article.

Algorithm 3 Building Network

Input: Preprocessed CSV file
Output: A network of article-indices referencing base article
1: *n_hop* ++
if *Ind_CitedBy[ind] is 0* **then**
  return NULL
end if
2: d={}
3: **for** *i in range(Ind_CitedBy[ind])* **do**
  if *Visited[ind][i] is NULL* **then**
    Visited[ind][i]=n_hops d[CitedBy[ind][i]]=go to step 1
  end if
  if *n_hops<=Visited[ind][i]* **then**
    d[CitedBy[ind][i]]=go to step 1
  end if
  if *d is empty* **then**
    loop
  end if
end for
5: Use *CitedBy* as net.

Algorithm 4 Diffusibility α

Input: Graph G
Output: Diffusibility of each node in the graph
1:Define reachability R of each node as
for *Each node in graph G* **do**
  Find the nodes which are reachable from the current node using DFS
end for
2: Calculate α.
for *each row in R* **do**
  for *each column in R* **do**
    if *R[row][column] > 0* **then**
      α[row] +=1
    end if
  end for
end for
3:**for** *Each i in* α **do**
  if $\alpha[i] == 0$ **then**
    α[i] = 1 to assign some diffusibility even to leaf nodes.
  end if
end for

claim from the same data-driven standpoint upon which the influence model is based, and we contend that it is true.

The technique of establishing the cause-and-effect link between two systems using their time-series measurements is known as causality testing. As important as determining the direction of causality is the degree of influence from the cause to the effect (assuming one exists). In cognitive neurology, climatology, physics, econometrics, social sciences, and engineering, causality testing has a wide range of applications [30]. We employed Wiener-Granger Causality or G-causality [31] and Transfer Entropy [32], [33], [34] as common model-free/data-driven techniques.

Causal analysis is performed at the article level on the H-index of the journal and the influence diffusion score to see the impact of individual article metrics on the diffusion score. Transfer Entropy is used for this causality study. To do so, 264 articles published by 18 institutions from the corpus are analyzed. For joint probability distributions, we used the MuTE toolbox [35] and for the binning estimator, non-uniform embedding is used. Set up of 6 bins and 5 past lags are considered. 100 surrogates were picked at a significance level of 0.05. The TE value between the H-index value of the journal in which article is published and the influence score is 0.102 bits, according to our findings. This implies that causal information flows from these two entities, supporting the hypothesis that publishing in good research forums improves the diffusion of influence of the articles and, as a result, institutions.

IV. CONCLUSION

In this paper, a novel influence diffusion model to analyze the influence spread of the research institutions in a domain is proposed, which goes beyond the volume of publications and citation count and is only based on the quality of research output of the institutions. In the bottom-up approach, we quantified the diffusion of the articles published by the institution and then computed diffusion scores for the institutions. We build a diffusion model that is reachability-based and does not consider positional/centrality parameters of the nodes in the network since these are used to identify and rank important/influential nodes. The proposed

Algorithm 5 Computation of Parameter β

Input: Graph G
Output: Value of β of each node in the graph
1: **for** *each row in G* **do**
  for *each column in G* **do**
    dout[row] -= g[row,column] outgoing edges
    din[row] += g[column,row] incoming edges
  end for
  d[row] = dout[row] + din[row]
end for

TABLE 6. Capturing diversity in publication type, the table highlights the university-wise normalized scores for articles published in conference proceedings. The total number of citation networks analyzed are 646.

Univ	Articles	Tol. no. of Direct citations	Score
UN1	61	468	458.10
UN2	12	80	31.64
UN3	56	568	355.22
UN4	262	2605	1877.23
UN5	80	604	598.86
UN6	49	361	289.41
UN7	1	2	3.38
UN8	13	58	112.06
UN9	6	22	26.59
UN10	9	66	97.01
UN11	6	53	28.99
UN12	11	39	29.71
UN13	17	109	57.50
UN14	5	25	24.29
UN15	20	179	171.61
UN16	14	76	49.14
UN17	8	81	31.32
UN18	16	219	162.96

TABLE 7. Capturing diversity in Publication type, the table highlights the university-wise normalized scores for articles published in Journals. The total number of citation networks analyzed are 264.

Univ	Tot. no. of Articles	Tol. no. of Direct Citations	Score
UN1	22	142	226.36
UN2	12	94	41.36
UN3	35	402	184.85
UN4	60	1122	554.62
UN5	27	211	131.45
UN6	32	370	331.18
UN7	3	6	5.16
UN8	4	20	34.32
UN9	3	35	22.45
UN10	10	73	52.81
UN11	15	52	69.99
UN12	3	4	3.48
UN13	4	43	17.53
UN14	1	5	1.51
UN15	14	173	114.18
UN16	4	22	18.11
UN17	8	69	25.97
UN18	7	128	98.58

TABLE 8. Capturing diversity in publication type: the table highlights the university-wise normalized scores for articles published in journals and conferences. The total number of citation networks analyzed is 910.

Univ	Total no. of Articles	Tol. no. of Direct Citations	Score
UN1	83	610	621.90
UN2	24	174	63.70
UN3	91	970	573.16
UN4	322	3727	2309.74
UN5	107	815	794.51
UN6	81	731	485.16
UN7	4	8	11.05
UN8	17	78	145.24
UN9	9	57	40.29
UN10	19	139	198.54
UN11	21	105	101.70
UN12	14	43	36.82
UN13	21	152	71.23
UN14	6	30	28.84
UN15	34	352	290.77
UN16	18	98	62.70
UN17	16	150	62.19
UN18	23	347	235.53

TABLE 9. Capturing diversity in publication type, the table highlights the university-wise SNIP normalized scores for the articles published in journals. The total number of citation networks analyzed is 264.

Univ	Total no. of Articles	Tol no. of Direct Citations	Score
UN1	22	142	248.55
UN2	12	94	51.08
UN3	35	402	178.61
UN4	60	1122	572.49
UN5	27	211	116.20
UN6	32	370	344.60
UN7	3	6	4.85
UN8	4	20	33.38
UN9	3	35	18.51
UN10	10	73	48.53
UN11	15	52	75.65
UN12	3	4	4.20
UN13	4	43	16.27
UN14	1	5	1.51
UN15	14	173	115.67
UN16	4	22	16.83
UN17	8	69	31.52
UN18	7	128	97.82

model analyzes the influence spread in a citation network and hence consider nodes in the single-hop as well as in multiple hops from the source node. The amount of influence diffused depends on the position of the node relative to the source node and the in-degree and out-degree of the nodes. The heat kernel is modified suitably to reflect these considerations and is used to compute the amount of influence diffused. A salient feature of the proposed approach is that the high diffusion scores do not just reflect the high volume of publications or citation count. We establish that influence diffusion does not depend only on the number of publications and received citation count but the outreach that each article has – and this needs to be analyzed in its citation network, so direct and indirectly influenced articles are considered in this work. The results obtained are in line with the proposed hypothesis. From this same data-driven frame of reference, we performed a cause-effect investigation using causality testing and argued that publishing in a conference or journal with a good h-index or SNIP *causes* scholarly influence. The proposed model can be effectively extended to be used in cases where the influence diffusion of the research entity has to be quantified. The highly influential articles, the ranking of researchers and institutions with 'highly influential papers', performance evaluation of the research output of these entities are some examples where we envisage the use of our method.

V. FUTURE WORK AND SCALABILITY CONSIDERATIONS

Given the focus of the work on theoretical formulation and initial proof-of-concept analysis, the proposed diffusion framework models influence diffusion in citation networks efficiently. Considering the constantly increasing volume of scholarly publications, we acknowledge that real-world scalability depends on computing efficiency. The two main steps of our algorithms are: Citation network construction from the metadata, and Diffusion Score computation using reachability and influence measures. The worst-case time complexity for dense graphs may result from operations like reverse mapping, adjacency matrix creation, graph traversals, and Graph Laplacian operations. As a part of future optimizations, DFS with early pruning can be incorporated to avoid redundant traversals of the graph nodes, which can help bring down time complexity further and improve practical efficiency. For future large-scale deployments, we also propose potential improvements as: Replacing the dense adjacency matrix with sparse representations and exploring parallel ang GPU-accelerated operations.

APPENDIX

A. ALGORITHMS DEPICTING THE STEPS TAKEN TO CREATE CITATION NETWORKS

Algorithms 2–5.

B. TABLES TO CAPTURE DIVERSITY IN PUBLICATION TYPE

Tables 6–9.

REFERENCES

[1] P. S. Aithal and P. Kumar, "ABC model of research productivity and higher educational institutional ranking," *Int. J. Educ. Manage. Eng.*, vol. 6, no. 6, pp. 74–84, Nov. 2016.

[2] J. K. Pal and S. Sarkar, "Evaluation of institutional research productivity," *DESIDOC J. Library Inf. Technol.*, vol. 40, no. 1, pp. 406–417, Feb. 2020.

[3] D. Docampo and J.-J. Bessoule, "A new approach to the analysis and evaluation of the research output of countries and institutions," *Scientometrics*, vol. 119, no. 2, pp. 1207–1225, May 2019, doi: 10.1007/s11192-019-03089-w.

[4] F. A. Massucci and D. Docampo, "Measuring the academic reputation through citation networks via PageRank," *J. Informetrics*, vol. 13, no. 1, pp. 185–201, Feb. 2019.

[5] H. Cao, C. Gao, and Z. Wang, "Ranking academic institutions by means of institution–publication networks," *Phys. A, Stat. Mech. Appl.*, vol. 629, Jan. 2023, Art. no. 129075, doi: 10.1016/j.physa.2023.129075.

[6] M. Wang, X. Zhang, H. Zhong, and D. Wang, "AIRank: An algorithm on evaluating the academic influence of papers based on heterogeneous academic network," *J. Inf. Sci.*, Feb. 2023.

[7] G. Abramo, "Revisiting the scientometric conceptualization of impact and its measurement," *J. Informetrics*, vol. 12, no. 3, pp. 590–597, Aug. 2018.

[8] S. R. Dey, A. Mathur, B. S. Dayasagar, and S. Saha, "ALIS: A novel metric in lineage-independent evaluation of scholars," *J. Inf. Sci.*, vol. 49, no. 4, pp. 1007–1030, 2023.

[9] X. Bai, F. Zhang, J. Ni, L. Shi, and I. Lee, "Measure the impact of institution and paper via institution-citation network," *IEEE Access*, vol. 8, pp. 17548–17555, 2020, doi: 10.1109/ACCESS.2020.2968459.

[10] A. Zareie, A. Sheikhahmadi, and M. Jalili, "Influential node ranking in social networks based on neighborhood diversity," *Future Gener. Comput. Syst.*, vol. 94, pp. 120–129, May 2019.

[11] K. K. Abdulmajeed, "Influential nodes identification in complex networks: Sampling approach," *Int. J. Intell. Syst. Appl. Eng.*, vol. 12, no. 21, pp. 1907–1916, 2024.

[12] K. Berahmand, E. Nasiri, S. Forouzandeh, and Y. Li, "A preference random walk algorithm for link prediction through mutual influence nodes in complex networks," *J. King Saud Univ. Comput. Inf. Sci.*, vol. 34, no. 8, pp. 5375–5387, Sep. 2022.

[13] P. Scarabaggio, C. Paolo, and D. Raffaele, "Mariagrazia A novel approach for fast and effective influence spread maximization in social networks," *TechRxiv. Preprint*, Jan. 2020, doi: 10.36227/techrxiv.13153139.v1.

[14] Z. Li, W. Piao, Z. Sun, L. Wang, X. Wang, and W. Li, "User real-time influence ranking algorithm of social networks considering interactivity and topicality," *Entropy*, vol. 25, no. 6, p. 926, Jun. 2023, doi: 10.3390/e25060926.

[15] S. R. Dey, S. Kotian, A. Agarwal, A. Lalan, G. S. Sampatrao, and S. Saha, "Epigraphiology: A hybrid approach for measuring and analyzing influence diffusion in article networks," *J. Scientometric Res.*, vol. 13, no. 2, pp. 615–624, Aug. 2024, doi: 10.5530/jscires.13.2.48.

[16] H. Cui, A. Zeng, Y. Fan, and Z. Di, "Identifying the key reference of a scientific publication," *J. Syst. Sci. Syst. Eng.*, vol. 29, no. 4, pp. 429–439, Aug. 2020, doi: 10.1007/s11518-020-5455-3.

[17] V. Aman and J. Gläser, "Investigating knowledge flows in scientific communities: The potential of bibliometric methods," *Minerva*, vol. 63, pp. 155–182, Oct. 2024, doi: 10.1007/s11024-024-09542-2.

[18] P. Sarkar, G. S. Sampatrao, and S. Saha, "NLRIS: Modeling and analysis of non-local influence of research output of institutions," *J. Scientometric Res.*, vol. 11, no. 1, pp. 79–88, Jun. 2022.

[19] M. Wang, J. Zhang, S. Jiao, and T. Zhang, "Evaluating the impact of citations of articles based on knowledge flow patterns hidden in the citations," *PLoS ONE*, vol. 14, no. 11, Nov. 2019, Art. no. e0225276.

[20] J. Ravenscroft, M. Liakata, A. Clare, and D. Duma, "Measuring scientific impact beyond academia: An assessment of existing impact metrics and proposed improvements," *PLoS ONE*, vol. 12, no. 3, Mar. 2017, Art. no. e0173152.

[21] S. L. De Groote, M. Shultz, and N. R. Smalheiser, "Examining the impact of the national institutes of health public access policy on the citation rates of journal articles," *PLoS ONE*, vol. 10, no. 10, Oct. 2015, Art. no. e0139951.

[22] 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.

[23] L. Egghe, "Theory and practise of the g-index," *Scientometrics*, vol. 69, no. 1, pp. 131–152, Oct. 2006.

[24] B. Jin, L. Liang, R. Rousseau, and L. Egghe, "The R- and AR-indices: Complementing the h-index," *Chin. Sci. Bull.*, vol. 52, no. 6, pp. 855–863, Mar. 2007.

[25] J. Yang and Z. Liu, "The effect of citation behaviour on knowledge diffusion and intellectual structure," *J. Informetrics*, vol. 16, no. 1, Feb. 2022, Art. no. 101225.

[26] X. Bai, F. Zhang, J. Hou, I. Lee, X. Kong, A. Tolba, and F. Xia, "Quantifying the impact of scholarly papers based on higher-order weighted citations," *PLoS ONE*, vol. 13, no. 3, Mar. 2018, Art. no. e0193192.

[27] J. Li and P. Willett, "ArticleRank: A PageRank-based alternative to numbers of citations for analysing citation networks," in *Proc. Aslib*, vol. 61, 2009, pp. 605–618.

[28] M. Doo and L. Liu, "Extracting top-k most influential nodes by activity analysis," in *Proc. IEEE 15th Int. Conf. Inf. Reuse Integr. (IEEE IRI)*, Aug. 2014, pp. 227–236.

[29] M. Paidar, S. Seddighi, and A. Harounabadi, "Identifying top-k most influential nodes by using the topological diffusion models in the complex networks," *Int. J. Adv. Comput. Sci. Appl.*, vol. 8, no. 6, 2017.

[30] A. Kathpalia and N. Nagaraj, "Measuring causality: The science of cause and effect," *Resonance*, vol. 26, no. 2, pp. 191–210, Feb. 2021.

[31] C. W. J. Granger, "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, vol. 37, no. 3, pp. 424–438, Aug. 1969.

[32] T. Schreiber, "Measuring information transfer," *Phys. Rev. Lett.*, vol. 85, no. 2, pp. 461–464, Jul. 2000.

[33] R. Vicente, M. Wibral, M. Lindner, and G. Pipa, "Transfer entropy—A model-free measure of effective connectivity for the neurosciences," *J. Comput. Neurosci.*, vol. 30, no. 1, pp. 45–67, Feb. 2011.

[34] L. Barnett, A. B. Barrett, and A. K. Seth, "Granger causality and transfer entropy are equivalent for Gaussian variables," *Phys. Rev. Lett.*, vol. 103, no. 23, Dec. 2009, Art. no. 238701.

[35] A. Montalto, L. Faes, and D. Marinazzo, "MuTE: A MATLAB toolbox to compare established and novel estimators of the multivariate transfer entropy," *PLoS ONE*, vol. 9, no. 10, Oct. 2014, Art. no. e109462.

GAMBHIRE SWATI SAMPATRAO received the Ph.D. degree in computer and information sciences from VTU, Karnataka, in 2024. She is currently an Associate Professor with the CSE Department, PES University, Bengaluru, India. Her research interests include data analytics, machine learning, and social network analysis.

ANISHA KHETAN received the bachelor's degree in computer science from PES University, EC Campus, Bengaluru, India, in 2020, and the M.S. degree in data science from the University of Minnesota, Twin Cities, in 2023.

NITHIN NAGARAJ (Senior Member, IEEE) received the bachelor's degree in electrical and electronics engineering from the NITK, Surathkal, in 1999, the master's degree in electrical engineering from the Rensselaer Polytechnic Institute (RPI), Troy, NY, USA, and the Ph.D. degree from the National Institute of Advanced Studies (NIAS), Bengaluru, in 2010. He is currently a Professor with the National Institute of Advanced Studies, Indian Institute of Science Campus.

SNEHANSHU SAHA (Senior Member, IEEE) received the Ph.D. degree in applied mathematics from The University of Texas at Arlington. He is currently a Professor of computer science and the Head of the AI Research Center, APPCAIR, Birla Institute of Technology and Science, Pilani (BITS Pilani), K. K. Birla Goa Campus. His research interests include the theory of deep learning, computational learning, and optimization.

SUDEEPA ROY DEY is currently an Associate Professor with the Department of CSE, PES University, EC Campus. Her Ph.D. is in the area of scientometrics and network mining and graph analytics.

ARCHANA MATHUR (Senior Member, IEEE) received the master's degree in engineering from Bengaluru University and the Ph.D. degree in machine learning and scientometrics. She has been a Research Assistant with Indian Statistical Institute, Bengaluru. She is currently a Professor with Nitte Meenakshi Institute of Technology. Her research interests include foundations of machine learning, deep learning, and optimization techniques.