

Epigraphiology: A Hybrid Approach for Measuring and Analyzing Influence Diffusion in Article Networks

Sudeepa Roy Dey¹, Shivani Kotian¹, Anmol Agarwal², Arshika Lalan², Gambhire Swati Sampatrao¹, Snehanshu Saha²

¹Department of CSE, PES University, Karnataka, INDIA.

²APPCAIR and CSIS, BITS Pilani K.K. Birla Goa Campus, Goa, INDIA.

ABSTRACT

Identifying influential nodes in an article network is crucial for understanding the dynamics of information propagation and its impact on various applications. Traditional methods often rely on citation-based analysis or network structure, overlooking the intricate dynamics of diffusion and node linkages. In this research, we propose a novel scoring model, named "Epigraphiology," which combines these aspects to compute and analyze the elements contributing to the spread of influence in article networks. To evaluate the effectiveness of our approach, we employ real published article networks with around 904 articles downloaded from the WOS (Web of Science) with total cited references of 32084 in the field of cloud computing from 2010 to 2015. By leveraging the SIR (Susceptible-Infected-Removed) model, we compare the dynamics of articles in the network with the transition of states, highlighting the diffusion process. Additionally, we derive the Reproduction number (R_0) for our model, serving as an indicator of the potential spread of influence. Our findings showcase the following key contributions: (a) Epigraphiology introduces a novel methodology for measuring the diffusion capacity of an article's influence in a hybrid manner, combining diffusion dynamics and node linkages. (b) Contrary to traditional approaches that primarily consider the number of citations (in degree), our results reveal that articles with lower citation counts can still act as super-spreaders, reflecting the ground-truth influence scores. Cross-validation of an article's influence diffusion score is performed, shedding light on the significant factors contributing to its spread within the network. By bridging the gap between diffusion dynamics, node linkages, and influence measurement, Epigraphiology offers a comprehensive approach to understanding and quantifying the spread of influence in article networks. This research holds implications for various fields and applications where the identification of influential spreaders is paramount in leveraging information dissemination and impact assessment.

Keywords: Graph network, Influence diffusion, Epigraphiology, Information Retrieval, Influential article mining, Citation-based social network.

Correspondence:

Sudeepa Roy Dey

Department of CSE, PES University,
Karnataka, INDIA.

Email: sudeepar@gmail.com

Received: 08-05-2023;

Revised: 04-07-2023;

Accepted: 05-07-2024.

INTRODUCTION

The ranking of an article node based on "influence" in a citation network is often based on the Centrality Measure (CM) and its variations. The common strategies used vary from graphical centrality measures such as degree, closeness, and betweenness, to diffusion-based methods, like Page-Rank, Leader-Rank, and epidemiological models.^[1] Authors A. Landher *et al.* explained how in Social Network Analysis (SNA) literature a wide range of CM exists to quantify the interlink of individual entities associated with the social network.^[2] These commitments from SNA literature permit the general conclusion that distinctive

CM frequently indicates various outcomes for the centrality of individual entities. Everett and S. P. Borgatt discussed the limitations of CM and explained three ways to further improve the basic concept of centrality.^[3] In the first method, the centrality is applied to groups as well as individuals. In the second level, two-mode data is used that applies all the tools and concepts of centrality to this data set. In the third method, centrality is applied to the innermost and external periphery structure of a network. The authors Zeng and Zhang proposed a Mixed Degree Decomposition (MDD) procedure based on the K-shell decomposition method. The MDD approach is shown to outperform the existing ranking approaches that are based on degree methods.^[4] They also introduced some interesting additions to the existing concepts of degree, closeness, and betweenness centrality as distinguished by.^[5,6] Chen, Gao, *et al.* proposed the local ranking algorithm named ClusterRank which considers not only the number of neighbors and the neighbors'



DOI: 10.5530/jscires.13.2.48

Copyright Information :

Copyright Author (s) 2024 Distributed under
Creative Commons CC-BY 4.0

Publishing Partner : EManuscript Tech. [www.emanuscript.in]

influences but also the clustering coefficient.^[7] The authors' Zhu *et al.* proposed a unique approach to ranking individual nodes of a real-world communication network based on their roles in such diffusion processes.^[8,9] The existing evaluative methods explore the citation network of an article and try to trace the diffusion path using indirect citations or by exploring the similarity index.^[10] Also, the existing Article Influence Score (AIS) as reported by the SCI-Web of Science is an average influence of a journal's article for the first five years after publication. These metrics do not consider the spreading abilities of an article node. However, a lot of literature exists on the SIR spreading model. It is used for simulating the spreading processes in networks to evaluate the performance of the algorithm as explained in.^[11] Nevertheless, scientific literature should place some emphasis on the strength and spread of influence propagation of a node in a short time while scoring the "Influentiality" of a node.

In this paper, we propose a hybrid approach called 'Epigraphiology' to identify the influential nodes in a citation network using the SIR model. A.G. McKendrick and W.O. Kermack formalized the famous modern mathematical epidemic model named the Susceptible-Infected-Recovered (SIR) compartmental model^[12] when studying the spreading pattern of plague. The mathematical model for the spread of infection was explained in a series of works by.^[13] The concept of reproduction number was first introduced by^[14] where it was shown through clinical trials that a threshold of the mosquito population below a critical level would be sufficient to control malaria. These compartmental models are found to be useful in estimating the diffusion or spread of an infection in a susceptible population. We have harnessed the reproduction number R_0 as the key indicator of the spread of influence diffusion across citation networks.

The value of R_0 is derived mathematically using the graphical parameters from a real-world citation network of articles. Further weights are assigned to each of the features and an optimized score is calculated which reflects the Influence of an article in a network based on the direct(primary) and indirect citations(secondary infections).

Objective and Problem Statement

The current scholastic evaluation methods lack a metric that can effectively capture an article's influence diffusion within a domain, independent of the ecosystem it grows in. Existing graph-based metrics, such as K-core and centrality metrics, do not adequately reflect the article's "influentiality".^[15] This poses a challenge for agencies seeking to evaluate a group of researchers' work and for top-tier universities looking for evidence of influential research capability during faculty recruitment or funding grants. Additionally, there is a need for a metric that funding/grant/patent organizations can employ to assess the impact their funding had on a funded project. Therefore, the main problem addressed in

this study is the absence of a comprehensive metric that purely quantifies an article's influence-spreading abilities, goes beyond citation count to measure impact, and allows for evaluating the impact of funding on a project. The research aims to develop a unique graph-based method that combines the positional and diffusion capabilities of nodes in the academic network. The proposed algorithm was tested using real data and then compared with existing methods to draw significant inferences and validate its effectiveness in measuring and evaluating academic influence.

We address the following issues during our experimental study

Is there any metric other than the existing ones that will quantify the article purely based on its influence-spreading abilities in a domain?

Is there any metric that goes beyond citation count and quantifies an article's impact?

Is there a metric that funding/grant/patent organizations can use to evaluate the impact the funding had on a funded project?

The rest of the paper is organized as follows. The "Methods" identified for Epigraphiology are

presented in the following section. The Section "Experimental setup" discusses the complete setup for the experiment with the construction of the dataset and construction of the network with the derivation of the parameters and finally the influence score computation. The "results and discussion" section is written to provide insight into the algorithm and its highlights.

METHODOLOGY

The methodology for deriving the R_0 for our model and developing the unique graph-based method to measure academic influence involved the following steps:

Data set construction

A data set with around 904 articles downloaded from the WOS (Web of Science) with total cited references of 32084 was constructed. Next, a network representation was built using the collected data. Nodes in the network represented academic articles or researchers, and edges represented relationships such as citations. Table 1 list outs the Table search string and Table 2 is a snapshot of the dataset

Table 1: Fields searched.

Term	Fields extracted
Nodes	Article ID, Authors name, year, volume
Edges	Citation count
Indegree	Citation recieved
Outdegree	Cited references

Table 2: Snapshot of article dataset.

Id	Label	indegree	outdegree	Degree
17	Zissis, 2012, V28, P583	47	0	47
65	Wang, 2011, V22, P847	38	0	38
153	Grobauer, 2011, V9, P50	23	0	23
126	Ren, 2012, V16, P69	23	0	23
87	Wang, 2012, V5, P220	22	3	25
64	Wang, 2013, V62, P362	17	2	19
103	Li, 2013, V24, P131	21	0	21
228	Lombardi, 2011, V34, P1113	15	0	15
158	Xiao, 2013, V15, P843	14	2	16
222	Rong, 2013, V39, P47	14	0	14
189	Wan, 2012, V7, P743	13	0	13
258	Hao, 2011, V23, P1432	13	0	13

Factor Identification

The obtained network was analyzed and various factors that influence the spread of influence within the domain were identified. This involved examining the network structure, node characteristics, influence transmission mechanisms and time dynamics.

Model Inception

Developing a mathematical model using an Epigraphiology algorithm that incorporates the identified factors to quantify the influence-spreading abilities of academic articles or researchers. The model was designed to capture the positional and diffusion capabilities of nodes in the network.

R0 Derivation

The developed model was further used to estimate the basic reproduction number (R0) for the academic network. R0 represents the average number of secondary influence transmissions caused by a single influential node. This estimation will provide an objective measure of the influence-spreading abilities within the domain.

Validation and Comparison

Validate the derived R0 value and the proposed graph-based method by comparing the results with existing methods or metrics. Conduct statistical analyses and draw appropriate inferences to demonstrate the effectiveness and uniqueness of the developed approach.

By following this methodology, we could address the research objectives and problem statement, develop a unique graph-based method for measuring academic influence, and provide valuable insights into evaluating influential research credentials within a domain.

Graphical Methods for Network Structure and Node Characteristics

The underlying structure of Epigraphiology is based on the graph which is created from the articles as nodes and edges representing citations. The first part of the algorithm is used to compute the positional strength of an article node in a graph. Our graph $G(V, E)$ had 'N' vertices which were the articles in a domain published within the years 2011-2015. The 'E' edges represented citations from one article to another. Next, the positional strength was computed using the following graphical measures as defined below:-

The In-degree of a node

In-degree is defined as a reflection of a node's importance in a citation network. This is because the directed edge from article 'i' to 'j' indicates that citing article 'i' is influenced/infected by the cited article 'j'.

Closeness Centrality

It is one of the most commonly used measures in citation networks, where nodes with high closeness centrality explain the highly favorable position of the node to spread the influence around.

Betweenness Centrality

The Betweenness centrality method calculates the amount of influence a node has over the information flow in a graph. It is often used to establish a connection between one part of a graph to another.

Eigen Vector Centrality

It is an extremely important graphical measure where the nodes with higher in-degree have a high score. So a connection from

such high-scoring nodes is considered more important than low-scoring ones.

The SIR-Based-Influence Spreader Model

One of the efficient methods to quantify the influence of an article node in a network is by investigating its trajectory path. The Susceptible-Infected-Removed model (SIR Model) is one of the commonly used methods which can simulate the spread through compartmental stages. The SIR model is embraced in "Epigraphiology" to assess the spreading capacity of a node. We derive the formula for Reproduction number (R_0) as per our model and track the secondary infections of an article. Each node in the network represents an article published in a year and labeled as Susceptible (S). The citation to each of these articles creates a transition from S to I (Infected). Often articles do not get any citations even after a period of five years citation window. These articles are then moved from S to R (Removed). For further analysis, the articles are classified into three groups based on their position in our citation window from 2011-2015:

Highly Susceptible

Those articles which are published in 2011-12 and hence are highly susceptible to maximum infections (citations) owing to maximum exposure in a domain for 5 or 4 years.

Moderately Susceptible

Those articles were published in 2013 and are aged 3-4 years. These articles are moderately exposed to infections and may be capable of becoming efficient spreaders in the future.

Nascent

Those articles published towards the end of our citation window with just existence of 2 years and yet may have the potential of being a super spreader.

Mathematical Modelling with Time Dynamics

In the citation network context, $S(t)$ is the number of articles published in the domain of "security issues in cloud computing" at a time 't' which is the year of publication. $I(t)$ is the number of articles that get citations from other articles representing the spread of infection. $R(t)$ is the number of articles that do not get citations. The SIR model has certain parameters like infection rate (α), Transmissibility rate (β), and recovery rate (γ) which are crucial to reflect the dynamic nature of a complex network. Here α is the rate at which a node gets infected in a population of susceptible nodes while β is the average rate of contact between susceptible and infected individuals and γ is the rate of removal from the population. Table 3 lists out the analogies between the SIR model and our model.

We represent the states as the proportions of the network as follows.

$$S = \frac{S(t)}{N}, I = \frac{I(t)}{N}, R = \frac{R(t)}{N} \quad (1)$$

$N(t)=S(t)+I(t)+R(t)$ the total population size.

$$\frac{dS}{dt} = -\beta * S * I/N \quad (2)$$

$$\frac{dI}{dt} = \beta * S * I/N - \gamma * I \quad (3)$$

$$\frac{dR}{dt} = \gamma * I \quad (4)$$

Derivation of Reproduction Number

In the study of information diffusion models, the Epidemics model has been potentially found to be highly effective in implicit networks. R_0 is a crucial parameter in the study of infectious diseases and epidemiology.^[16] It represents the average number of secondary infections that can be generated by a single infected individual in a population that is entirely susceptible to the disease. The most important use of R_0 is determining the contagiousness and transmissibility of an infectious disease. This property is suitably used in our model to simulate and measure the spread of article influence. The basic reproduction number must be greater than 1, otherwise, influence propagation will die off, i.e., $R_0 > 1$. Figure 1 shows the propagation of infection for $R_0=2$. The node at layer 0 is capable of spreading the infection to at-least two nodes which further can infect two other individuals.

To derive the R_0 for our model, we need to identify and analyze the various factors that affect the spread of influence. These factors play a crucial role in determining the rate at which influence is transmitted from one node to another in the network. By understanding and quantifying these factors, we can estimate the basic Reproduction number (R_0), which represents the average number of secondary influence transmissions caused by a single influential node.

The spread of influence of an article depends on the average rate of citations, duration of the infection, and transmittable rate. We define R_0 for our model as follows:

$$R_0 \propto \left(\frac{\text{infection}}{\text{contact}}\right) \cdot \left(\frac{\text{contact}}{\text{time}}\right) \cdot \left(\frac{\text{time}}{\text{infection}}\right) \quad (5)$$

More specifically in our Model

$$R_0 = \alpha \cdot \beta \cdot d \quad (6)$$

where α is the probability of infection given contact between a susceptible and infected individual, β is the average rate of contact between susceptible and infected individuals, and d is the duration of infection. In this study, the above parameters are recalculated to derive the Reproduction number that is suitable for our citation network as follows:-

In graph $G(V, E)$ Infection Rate of a node n which belongs to the set vertices N is defined as:

Table 3: Analogy between Epidemiology and Epigraphiology.

Elements of SIR	Epidemiology	Epigraphiology
Susceptible S(t)	An initial population that is likely to get the disease.	Node as articles in a direct citation network that are published and susceptible to being cited by other article nodes.
Infective I(t)	A population that is infected by the disease can infect the susceptible population.	Articles nodes that have been cited by virtue of in-degree more than equal to one.
Recovered: R(t)	The infected population that recovers from a disease.	Cited articles that do not get citations after a while.
Infection probability: α	The rate at which the disease spreads.	The rate at which citations are received by an article is measured by a fraction of the in-degree of each node wrt to all nodes.
Transmissibility: β	The average number of times a person is infected in the duration of an epidemic.	Average citations received by an article per year
Removal rate: γ	The rate at which infected people recover.	The rate at which the article loses its influence.
Reproduction Number(R0)	An average number of secondary infections produced by any infected individual in the population.	The average number of nodes influence through indirect citations.

$$\alpha = \frac{\text{indegree}_n}{N}$$

In a graph G(V,E) Transmissibility Rate for a node n is defined as:

$$\beta = \frac{\sum_a^b \text{indegree}}{b - a}$$

where 'a' denotes the publication year of the article and 'b' is the end of the citation window. This can also be defined as the average number of citations received by an article per year.

The term 'd' indicates the duration of infection and in our model, this is calculated by investigating the years between the publication of the article and the maximum time frame which is 2015.

The removal rate (γ) is kept at 1 because ultimately all articles which are infected or susceptible state, tend to get removed as no citations are received by them after a certain period.

Epigraphiology Algorithm

The proposed algorithm detects the most influential nodes in a graph based on locality measures and spreading capabilities. The overall approach is shown in Figure 2.

The algorithm is a three-phased approach. In the first phase, the parameters are extracted from the citation network and graphical and epidemiology features are calculated. The value of the reproduction number is estimated from the α β values explained in the previous section. The next phase starts with assigning weights to the graphical features keeping R_0 as the target function. The extra tree classifier indicates the weights of each feature and finally the influence score is calculated as the weighted score. The algorithm below gives the complete details of the steps performed for influence-score computation.

Algorithm 1

Input: Nodes representing articles.

Edges representing Direct citation from a citing article to cited article.

Output: Influence Scores of each Article involved in the citation network .

Article nodes []=array to store article nodes in a network.

For each node x of node 'a' **do**

If x has cited a **then**

Append x to in-degree nodes []

end if

end for

for each node x in in-degree nodes [] **do**

Compute BC, CC, EC as ATTR VALUE

end for

for each node x in article nodes [] **do**

$R = ((\text{in-degree})/N) * \text{Avg edges} * D$

ATTR VALUE * W

end for

Experimental Setup and Results

In this section, we discuss the steps incorporated in creating the data set and the network. The overall methodology adopted can be explained in the following subsections.

Data Set and Graph Construction

The data set is created by downloading all articles published in the domain of "Security issues in cloud computing" from the Web of Science. To construct the network for "Epigraphology", all articles with at least indegree=1 or outdegree=1 both are

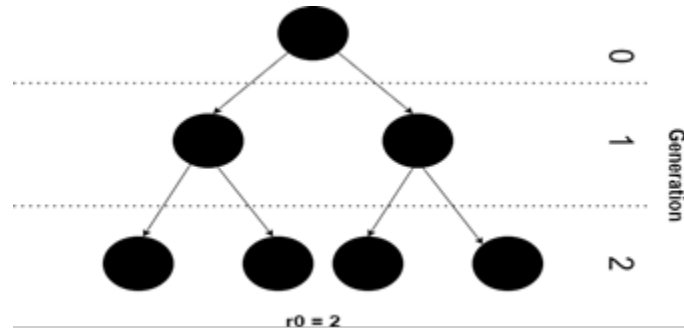


Figure 1: Initial spread of R_0 .

considered which means the article must get cited or cite one or more articles. This resultant network is a network with 460 nodes with 739 edges.

Figure 3 is an overview of our network. Once the graph is ready we start calculating the various local parameters for each node using the Python package networkx. The in-degree, out-degree, Closeness (CC), Betweenness (BC), and Eigen-Centrality (EC) are computed as discussed in the "Graphical Methods" section and stored in a data set along with the article ID and authors list. The next step is to calculate the R_0 as discussed in the "Epigraphology Algorithm" section. The parameters α and β along with the value of 'd' are listed below. The Reproduction number is one of the key predictors as it has the power of expressiveness to demonstrate the spread of influence. Table 4 represents the value of R_0 10 random articles. Interestingly highlighted node 379 indicates a good R_0 value even though the value α is less because of the very

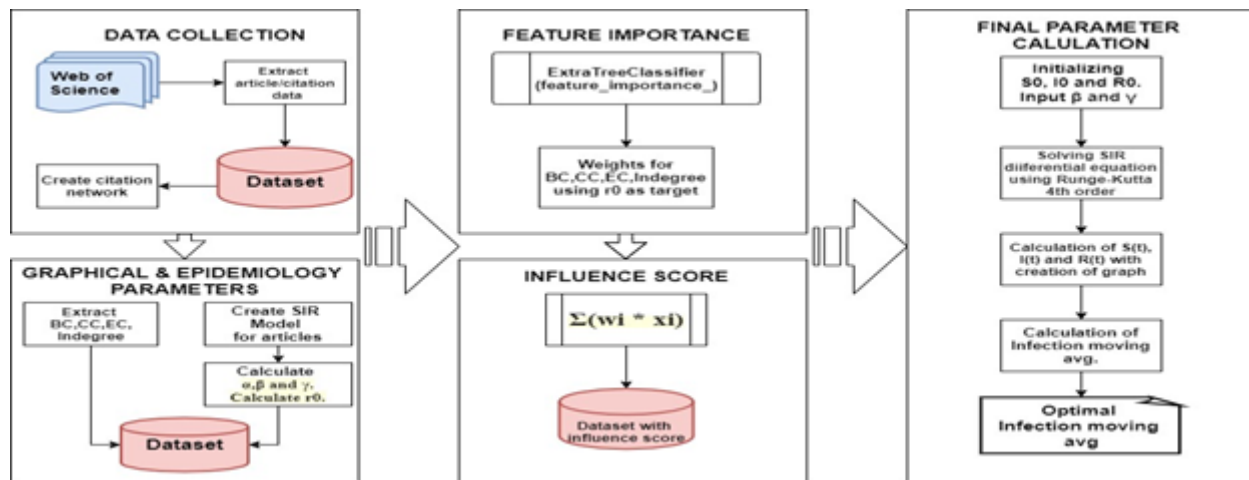


Figure 2: Overall Approach Adopted.

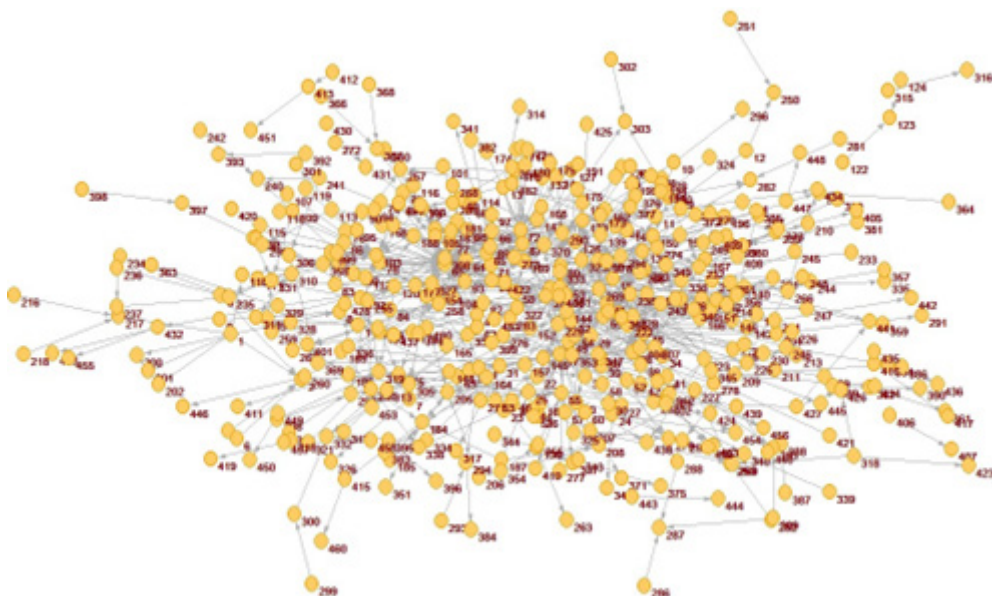


Figure 3: Final Citation Network.

Table 4: Reproduction number for 10 random articles.

Id	Label	Infectionprob- α	Transmissibility- β	D	R0
17	Zissis, 2012, V28, P583	0.1	11.5	0.25	0.29375
65	Wang, 2011, V22, P847	0.08	7.6	0.2	0.125565
379	Wei LF, 2014	0.039	9	0.5	0.176
153	Grobauer, 2011, V9, P50	0.05	4.8	0.2	0.048
126	Ren, 2012, V16, P69	0.05	5.75	0.25	0.071875
87	Wang, 2012, V5, P220	0.05	5.5	0.25	0.065761
64	Wang, 2013, V62, P362	0.04	9	0.333333	0.11087
103	Li, 2013, V24, P131	0.05	7	0.333333	0.106522
228	Lombardi, 2011, V34, P1113	0.03	3.2	0.2	0.02087
158	Xiao, 2013, V15, P843	0.03	4.667	0.333333	0.047346
222	Rong, 2013, V39, P47	0.03	4.667	0.333333	0.047346

Table 5: Comparison between K-core and R0 of a few articles.

Label	R0	Indegree	Paper Id	Corenum
Zissis, 2012, V28, P583	4.7	47	17	4
Wang, 2011, V22, P847	3.13913	38	65	4
Ren, 2012, V16, P69	1.15	23	126	4
Li, 2013, V24, P131	0.958696	21	103	3

high transmissibility rate. This node is an excellent example of the nascent published article category as though it is published in 2014 it is a potential super-spreader.

K-shell Decomposition: Comparative study with R0

The k-shell method is a popular existing method to identify influential nodes in a network. However, it uses only global information such as betweenness, and allocates the same core numbers to many nodes. The k-core method is used for static networks which have a fixed structure. The k-core of a particular graph 'g1' is defined as the maximal sub-graph of 'g1' having a degree of at least k. The K-shell method was applied to this dataset as well and a comparative analysis with R0 is performed. Table 5 shows a large number of articles are assigned the same shell numbers leading to difficulty in ranking articles. As shown in Table 5 both articles 17 and 126 have the same core number 4 but R0 is different for both articles. Also since only global information is used for assigning shell numbers, many nodes which are nascent and influential are not captured by K-decomposition.

Article Influence Diffusion Score Computation

Once all the technical parameters are collected from the above-described methods, the influence score is computed. We use the R0 values as the target function and assign weights to all the parameters using ExtraTreeClassifier. This is a classifying method, but we have used it to get the importance of attributes. The Article Diffusion score is now calculated by taking the product of attribute weight and attribute value. Table 6 shows the

influence scores of the top 10 highly influential articles with the in degrees. The results obtained are in line with our hypothesis that the citation received cannot truly reflect the influence of an article. Article ID 64 (highlighted) does prove our point as even with in-degree 17 its Influence score is at par with its established peer articles.

Figure 4 demonstrates the spread of node 64 as an influential article in the network. Also, the peak of infection is shown alongside.

Influence-Diffusion Cross-validation

Multivariate linear regression was performed over the Influence score using the centrality measure variables from the citation network as the independent variables. Mathematically, eq. 7 as follows:

$$Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 \quad (7)$$

Where Y = Influence diffusion score, X_1 = Closeness Centrality, X_2 = betweenness Centrality, and X_3 = Eigen Centrality. The model had an adjusted R^2 value of 0.733 and the F statistic was found to be significant. The coefficient for 'eigen centrality' was found to be significant. Table 7 shows the results of the regression.

It can be seen from Table 7 that the eigen-centrality value has a huge impact on the Influence diffusion score. The coefficients for Closeness Centrality and betweenness Centrality are insignificant indicating they do not have a significant impact on the Influence-Diffusion score and these variables are dropped

Table 6: Influence score of the top 10 articles.

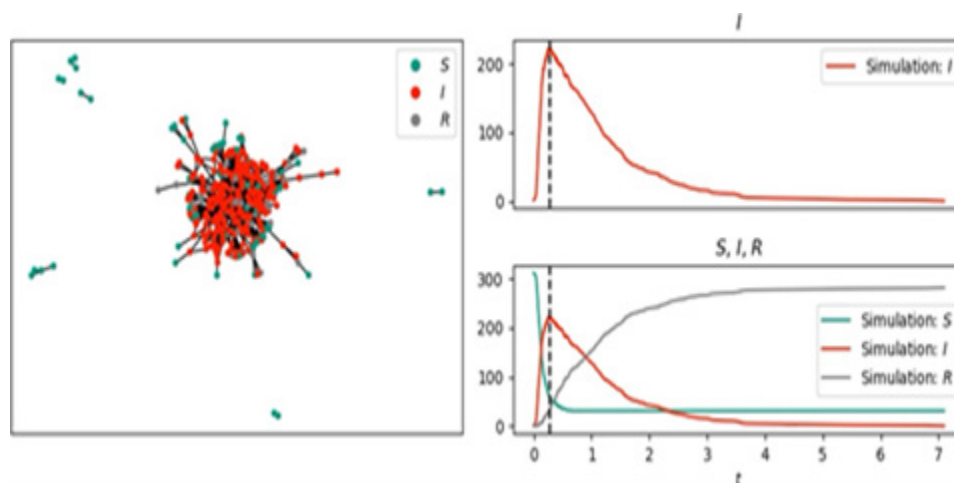
Id	Article	in-degree	Influence Score
17	Zissis, 2012, V28, P583	47	48.00977978
65	Wang, 2011, V22, P847	38	38.13913792
379	Wei LF,2014	18	32.5540227
64	Wang, 2013, V62, P362	17	32.43571049
103	Li, 2013, V24, P131	21	26.32993659
126	Ren, 2012, V16, P69	23	25.4006757
153	Grobauer, 2011, V9, P50	23	20.84784616
158	Xiao, 2013, V15, P843	14	20.72156364
222	Rong, 2013, V39, P47	14	17.6701442
48	Khan, 2013, V29, P1278	11	17.17718873

Table 7: Regression results for Influence Diffusion Score over the centrality measure variables.

Variable	Coefficient	Std. Error	t value	$p > t $
constant	9.3538	0.521	17.941	0.000
closenesscentrality	0.9248	0.647	1.428	0.158
betweenness centrality	0.7693	0.659	1.167	0.247
eigencentality	7.2417	0.537	13.495	0.000

Table 8: Regression results for the Influence diffusion Score over the Citation network variables.

Variable	Coefficient	Std. Error	t value	$p > t $
constant	9.3536	0.144	64.790	0.000
eigencentality	2.3362	0.270	8.663	0.000
beta	6.7014	0.255	26.257	0.000
gamma	0.8581	0.167	5.131	0.000

**Figure 4:** Final State of node 17 with the Highest Influence Score and infected nodes as red.

from the final model. The next goal is to detect multi-collinearity within the other independent variables. The Variable Inflation Factors (VIF) method was used to detect multi-collinearity among the independent variables. Variables In-degree and Infection probability had high VIF values and hence they were removed as well from the model. The model then ran a linear regression

using the citation network variables. Mathematically, the model can be represented by eq. 7 where Y = Influence diffusion score, X_1 = eigen centrality, X_2 = beta, and X_3 = gamma. All coefficients were found to be significant at the 95% level.

Table 8 shows the results of the regression. All variables play a significant impact in determining the Influence-diffusion score,

Table 9: Comparison with alternative methods

Parameters	AIS	Citation count	Epigraphiology
Scope	AIS evaluates the influence of entire journals based on the citations received by their articles over a five-year period.	Citation count measures the number of times an article has been cited by other scholarly works.	The proposed model focuses on evaluating influence diffusion of individual articles within a scholarly network.
Granularity	AIS provides a journal-level metric, aggregating the influence of all articles published in the journal.	It provides a quantitative measure of the impact of an article within the academic community but doesn't differentiate between direct and indirect influence or the spread of influence beyond citations.	The proposed model focuses on evaluating the influence diffusion of individual articles within a scholarly network.
Temporal Considerations	AIS is calculated based on citations accumulated over a five-year period, providing a snapshot of influence during that timeframe.	Citation counts typically reflect the cumulative number of citations received by an article over time. It doesn't necessarily differentiate between recent citations and those accumulated over a longer period.	The proposed model may capture the temporal dynamics of influence diffusion more directly, potentially accounting for both short-term and long-term effects as the influence spreads through the network.
Methodology	AIS is based on the Eigenfactor Score, which considers the network structure of citations and accounts for differences in citation patterns across disciplines.	Citation counts are straightforward to calculate and widely used as a proxy for research impact. However, they may be influenced by factors such as self-citations or citation practices within specific fields.	The proposed model utilizes graph network analysis to quantify the diffusion of influence from individual articles, considering factors such as direct and indirect connections between articles.
Application	AIS is commonly used to evaluate the impact and prestige of scholarly journals, aiding researchers, librarians, and funding agencies in decision-making.	Citation counts are commonly used by researchers, institutions, and funding agencies to assess the impact of scholarly publications. They play a crucial role in tenure and promotion decisions, funding allocations, and ranking of researchers and institutions.	The proposed model offers a systematic approach for evaluating individual scholarly contributions, aiding in the identification of influential articles and researchers within specific domains.

however, the beta value(Transmissibility Rate) seems to have the maximum impact. Thus, it can be verified from the above equations that the citation network variables have a significant impact on the Influence-Diffusion score, while the centrality measures do not have a huge impact.

Comparison of Epigraphiology with Alternative Methods

We compare the proposed Influence Diffusion model with the existing Article Influence Score (AIS) metric which evaluates the relative importance of a scholarly journal's articles within the citation network. It is calculated by dividing a journal's Eigenfactor Score by the number of articles published by the journal, normalized for differences in citation frequency across disciplines. We also compare our model with citation count as well on various parameters as shown in Table 9.

DISCUSSION

Epigraphiology is a unique method with graph graph-driven approach. We have performed various validations at each step to test Epigraphiology as an algorithm. 1) The use of R0 as one of the key predictors as it has the power of expressiveness to demonstrate the spread of influence is suitably compared with existing K-core decomposition in section 5.1.1. The results are discussed in detail to indicate the effectiveness of the same. 2) We Compared Epigraphiology with the existing metrics like Citation count and Article influence score in section 6.1. Our Model captures the spread of influence and the temporal dynamics of influence diffusion more directly, potentially accounting for both short-term and long-term effects as the influence spreads through the network. This is a unique quality that makes our model work on all real-life datasets. 3) Sensitivity analysis was performed by

cross-validating the influence diffusion score using Multivariate linear regression in section 6. Multivariate linear regression was performed over the Influence score using the centrality measure variables from the citation network as the independent variables. The results have been captured in Table 9 indicating Eigen Centrality as the major contributor.

CONCLUSION AND FUTURE SCOPE

This paper presents a graph network-driven model for computing the Influence Diffusion of scientific Articles. The proposed model was successfully applied to a real dataset of articles obtained from the Web of Science (WOS). The results were validated using regression analysis, highlighting the effectiveness and reliability of the model.

The ability to detect influential article nodes based on their spreading capabilities has significant implications for various domains, including ranking researchers and identifying novel inventions. Many universities prioritize researchers with highly influential papers, which go beyond traditional citation counts. The proposed model and quantitative measure offer a systematic approach for experts and selection committee members to identify significant contributions and potential promises in a specific domain.

The contributions of this research can be summarized as follows:

Introduction of a new influence diffusion model for evaluating scholarly contributions, demonstrated by computing Articles' Influence Diffusion scores using a real dataset of 904 articles.

Cross-validation of the model's results through regression analysis, ensuring its robustness and reliability.

Provision of empirical evidence for the proposed mathematical model by comparing its results with existing methods, establishing its consistency and effectiveness.

These contributions have the potential to yield important consequences. They can aid in the identification of articles suitable for national and international research awards, facilitate the determination of scholars' influence based on their scores, and trace institutional growth in non-opportunistic research by evaluating the fraction of faculty members with high Influence Diffusion scores. However the effectiveness of the model relies on the quality and completeness of the underlying dataset obtained from the Web of Science or similar sources. Inaccurate or incomplete data could lead to biased results and affect the

reliability of the Influence Diffusion scores. Hence data quality must be ensured before the model is applied for diffusion score computation.

Finally, the presented graph network-driven model provides a valuable tool for evaluating scholarly contributions and assessing the influence diffusion of articles. It offers a systematic and quantifiable approach that goes beyond traditional citation counts, enabling a more comprehensive understanding of research impact. By leveraging this model, institutions and decision-makers can make informed decisions about recognition, awards, and research evaluation, ultimately fostering the advancement of knowledge in various domains.

CONFLICT OF INTEREST

The authors declare that there is no conflict of interest.

REFERENCES

1. Dickman P. Diffusion tree restructuring for indirect reference counting. *ACM SIGPLAN Notices*. 2000;36(1):167-77.
2. Landherr A, Friedl B, Heidemann J. A critical review of centrality measures in social networks. *Business & Information Systems Engineering*. 2010;2:371-85.
3. Everett MG, Borgatti SP. Extending centrality. *Models and methods in social network analysis*. 2005;35(1):57-76.
4. Zeng A, Zhang CJ. Ranking spreaders by decomposing complex networks. *Physics Letters A*. 2013;377(14):1031-5.
5. Freeman LC. A set of measures of centrality based on betweenness. *Sociometry*. 1977;35-41.
6. Chen D, Lü L, Shang MS, Zhang YC, Zhou T. Identifying influential nodes in complex networks. *Physica a: Statistical mechanics and its applications*. 2012;391(4):1777-87.
7. Chen DB, Gao H, Lü L, Zhou T. Identifying influential nodes in large-scale directed networks: the role of clustering. *PloS one*. 2013;8(10):e77455.
8. Zhu W, Chen C, Allen RB. Analyzing the propagation of influence and concept evolution in enterprise social networks through centrality and latent semantic analysis. *Advances in Knowledge Discovery and Data Mining: 12th Pacific-Asia Conference, PAKDD 2008 Osaka, Japan, May 20-23, 2008 Proceedings 12 2008* (pp. 1090-8). Springer Berlin Heidelberg.
9. Shahzamal M, Jurdak R, Mans B, De Hoog F. Indirect interactions influence contact network structure and diffusion dynamics. *Royal Society open science*. 2019;6(8):190845.
10. Dickman P. Diffusion tree restructuring for indirect reference counting. *ACM SIGPLAN Notices*. 2000;36(1):167-77.
11. Yang FY, Li Y, Li WT, Wang ZC. Traveling Waves in a Nonlocal Dispersal Kermack-Mckendrick Epidemic Model. *Discrete & Continuous Dynamical Systems-Series B*. 2013;18(7).
12. Breda D, Diekmann O, De Graaf WF, Pugliese A, Vermiglio R. On the formulation of epidemic models (an appraisal of Kermack and McKendrick). *Journal of biological dynamics*. 2012;6(sup2):103-17.
13. Kermack WO, McKendrick AG. A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*. 1927;115(772):700-21
14. Ross SR. Malaria in India. *Nature*. 1911;88(2200):284-6.
15. Antonio Calì, Andrea Tagarelli, and Francesco Bonchi. 2020. Cores matter? An analysis of graph decomposition effects on influence maximization problems. In *Proceedings of the 12th AC Conference on Web Science (WebSci '20)*. Association for Computing Machinery, New York, NY, USA, 184-193. DOI:10.1145/3394231.3397908
16. Jones JH. Notes on R0. California: Department of Anthropological Sciences. 2007;323:1-9. Jones JH. Notes on R0. California: Department of Anthropological Sciences. 2007;323:1-9.

Cite this article: Dey SR, Kotian S, Agarwal A, Lalan A, Sampatrao GS, Saha S. Epigraphology: A Hybrid Approach for Measuring and Analyzing Influence Diffusion in Article Networks. *J Scientometric Res*. 2024;13(2):615-24.