

Recruitment Boosted Epidemiological Model for Qualitative Study of Scholastic Influence Network

Sudeepa Roy Dey^{1,*}, Snehanshu Saha², Rahul Reddy Nandyala³

¹PES University, Visvesvaraya Technological University, Belagavi, Karnataka, INDIA.

²CS&IS and Anuradha and Prashanth Palakurthi Centre for Artificial Intelligence Research (APPCAIR), BITS PILANI KK Birla Goa Campus, Goa, INDIA.

³PESIT-Bangalore South Campus, Electronic City, Bengaluru, Karnataka, INDIA.

ABSTRACT

Measuring the true influence of a researcher over the past few years has been an important problem in field of scientometrics as it not only facilitates funding organizations, academic departments, and researchers but also indicates the impact of scholarly influence. The existing author level ranking metrics such as *h-index*, measures of citation counts, can ignore much of the nuance and are also often criticised for being unfair owing to its purely quantitative approach. In this paper we propose an influence diffusion model based on the Epidemiological model variant called as Recruitment boosted SEIR. Our model tries to simulate the spread of infection with the growth of influence of a researcher by remodeling various existing parameters and building a new concept for qualitative study of prolific authors. Finally, the reproduction number is derived, and the scores are computed. To validate our influence diffusion model, we perform experiments on the real author dataset collected from web of science and compare the researchers influence with their paper's citation counts and *h-index*. Finally, we analyse the patterns about how researchers' influence ranking increased over time. Our studies also show the various changing patterns of researchers between different *h-index*.

Keywords: Epidemic, SEIR, Author citation, Author level Metrics, SIR model, SEIR model.

Correspondence

Sudeepa Roy Dey

PES University Visvesvaraya
Technological University, Belagavi,
Karnataka, INDIA.

Email id: sudeepar@gmail.com

Received: 23-04-2020

Revised: 14-10-2020

Accepted: 04-02-2021

DOI: 10.5530/jscores.10.1.13

INTRODUCTION

Capturing and measuring the true scholastic influence of an author/scholar's work has gained huge momentum owing to the breeding culture of "publish or perish" in most academic institutions. The major trade-off in this scenario is quality and relevance of the research work done. The existing bibliometrics ranking techniques are more focused on quantitative entities like citation counts, papers published etc without considering the impact and outreach of these research articles. A lot of work is being done in the field of author influence diffusion techniques using graphical measures like tree structures or graphs. In^[1,2] the authors have tried to create an influence diffusion tree that uses multiple-path asynchronous threshold (MAT) model for viral marketing in social networks. The authors have quantitatively measured influence and kept track of its spread and aggregation during the diffusion process. The MAT model captured both direct and indirect influence, depth-associated influence attenuation, temporal influence decay, and individual diffusion dynamics. Another major advancement in this domain has been by using the

epidemiological models. These infection models are found to be efficient in capturing the spread and impact of a disease.^[3] Mathematical modeling of infectious diseases is used to predict the transmission and the outcome of the diseases, which helps to provide possible counter measures to reduce the mortality rate or to eradicate the diseases. Inspired from such infection modeling, we have considered amoebiasis disease modeling to replicate the scientific growth and impact of a researcher. The populace or set of articles published by an author is partitioned into compartments, with the supposition that each person/article in a similar compartment has similar attributes. The SEIR model^[4] has an extended state as "exposed" which reflect the postponement between the obtaining of contamination and the infection state.^[5,6] The spread of infection is suitably interpreted as the influence spread of an author through his/her articles. The existing author level metrics like *h-index* and publication count indicate the quantitative aspect of researcher's growth.^[7] In this paper we have tried to compare these existing metrics with our influence scores using^[8] reproduction number. To study and compare the effect of the SEIR influence diffusion model, we have considered the various levels of authors and their growth trajectory. The following are the two categories namely:

1. Celebrity Authors- Researchers with higher *h-index* (greater than 50)
2. Rising star- Researchers with lower *h-index* (less than 10)

Copyright

© The Author(s). 2021 This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

The problem we try to address here is that of authors with more citation counts are called as influential authors, thus making the scholastic influence increasingly difficult to measure. The need to measure the true impact of scholarly work is achieved only when the true value of each citation is considered. We are aware that differentiating citations based on the source might seem highly conservative to many but by filtering the citations at the source, we intend to create an unbiased evaluation process. Our study is motivated by the existing gap in scholastic evaluation methods. We propose to discuss and address the following questions: -

- i. Can citation count alone capture the quality of an author's work? How can we use it for individual diffusion analysis?
- ii. Are all citations equal? Which citations indicate the global influence of an author's work?
- iii. How do we determine the impact of research to be considered important?
- iv. Can h -index and citation count truly justify the impact of an author?

The rest of the paper is organized into sections like motivation to this work followed by related work done so far in this field. The preceding sections discuss the author influence model in detail with data collection to data modeling, results and finally validation.

MOTIVATION

In recent years, the research community has started to explore different and innovative metrics that can define the "Scholastic Influence" for scientific and research assessments. Though citation analysis is one of the most popular technique used in scientometrics,^[9,10] this method involves counting the number of times a paper is cited. The underlying assumption is that the most influential researchers and their work will naturally receive more citations. A recent trend shows that many researchers have used these models to underplay the power of citation count by collaborating with peer groups or forming a network of known authors who coercively cite each other and boost their citations eventually.^[11] It is observed that there exists a lacuna in existing Author Level Metrics (AL Metrics) that measures the impact of an author by considering the citation counts such as h -index, i -10, etc.^[12] The main problem with these approaches is that the metrics are susceptible to artificial inflation of citation counts through community citation, self-citations, etc.^[6] that could cause the metrics to show incorrect or misleading data pertaining to the author's influence. This could lead to considering an author as substantially impactful or influential even though that may not be realistically true.^[13] With the use of the epidemiological

model the authors propose to overcome this drawback by considering other factors such as source of the citations received to make a better prediction of the influence of the author and make interesting revelations about the epidemic nature of the authors' influence.

RELATED WORK

A lot of researchers have tried to rank authors based on various metrics. In paper,^[14] the authors have used a rating approach that uses citations in a dynamic fashion, allocating rating by considering the relative position of two authors at the time of citation. The main objective of the paper was to introduce the notion of citation timing for relative ranking among the authors. The authors in paper^[4] have proposed a novel epidemic model, called as CISER model, for message propagation in DTN, based on Amoebiasis disease propagation in human population. The paper discusses in detail the various analogies of DTN and how the CISER model is instrumental in message propagation in many such networks. The authors have indicated the role of each compartment in the epidemiological model with derivations for each state using differential equations. In another paper^[15] the authors have implemented topic diffusion in web forums is modeled using the epidemiology model. The SIR model was adopted for the web forum. The model was evaluated on a large longitudinal dataset from the web forum of a major retail company and a dataset from a general political discussion forum. The experiment results revealed that the SIR model performed well in modeling topic diffusion in web forums.^[16] A lot of researchers have tried to work on scientometrics and in paper^[17] statistical characteristics of authors, co-authors, references, and citations are used to reveal the structure and dynamics of the research community and the intellectual environment of the field. The paper^[18] discusses scientometrics as quantitative (mainly statistical) study of any measurable aspect of scientific activity with the aim of understanding and, if possible, improving its operating mechanism. The authors have divided scientometrics as structural scientometrics whose purpose is the mapping of the structure of scientific communities, sets of documents, ideas, etc.^[19] Its typical techniques are, among others, graph theory, network analysis, cluster analysis.^[3] Whereas dynamic scientometrics purpose is to describe the space-time behavior of scientific information through scientometric objects (authors, publications, citations, etc.). Its typical methodological tools are ordinary and partial differential equations, stochastic models, and computer simulations.

PROBLEM STATEMENT AND OBJECTIVE

As already explored in the previous section, there are numerous metrics to measure the influence of an author in the public domain. However, these metrics rarely reflect influence of

a researcher's work. It is not beyond doubt that such metrics are effective universally, but a metric which is a descriptor of scholastic influence, would be handy for various agencies to evaluate researchers outside their peer groups and may further facilitate research assessment for outstanding credentials. We imply that an author may thrive by either receiving citations from pure strangers (aliens) or may depend on the support of the ecosystem of known collaborators. We would like to mention that in either case, we are not judgmental of the author or their citations. Nonetheless, there is a strong case for evaluating the author's influence spreading beyond known boundaries. This is a rare quality that science academics and agencies look for. We intend to bridge existing gaps in scholarly evaluation by addressing the following issues to be considered for our experimental study:

- i. Can we define a parameter given that field normalization has been done across research domains in computer science, minimizing the possibility of skewness/bias in citations that will compute the extent of outreach of research scholars?
- ii. Is there any way other than the existing ones that will determine the potential for independent research in early career of scientists?
- iii. Can we qualitatively measure the spread and growth of a researcher given the citation data?

We answer these questions positively, in the manuscript via theoretical model and empirical evidence. Citation count is often used as a metric to rank authors and their work. The authors propose another new approach as a natural consequence of such an exercise by tracing the source of citations and calculate the various parameters by deriving formulas and equations for the proposed epidemiological SEIR model. Epidemiological models are deterministic models that can be suitably modified to explore the dynamics of citations flow in an author's network. The epidemiological models have the following variants

- i. SIR model is an epidemiology model that studies the rate of infected people in a closed population over a fixed time. The model compartments three stages namely S as number susceptible, I as number infectious, and R as number recovered (immune).
- ii. SIS Model where there are two states only as susceptible and infected.
- iii. SEIR Model has an additional state called Exposed which indicates that the infection is local and still not in infectious state.

For our problem domain, we selected SEIR model to facilitate the citations movement based on source of citing articles.

Initially all articles published by an author is considered as susceptible. Based on the type of citation received the transition of articles begin from S to I or S to E. The removal of articles is based on no citations received for over a period of three years consecutively. The next section explains the model formulation for the same.

AUTHOR INFLUENCE MODEL FORMULATION

1. Data collection and Modeling

The complete methodology adopted is shown in Figure 1 below. One of the major tasks was to collect data of various authors and store them for further analysis. The data was collected from Web of science using the advanced search option for authors in field of computer science.

Next using the filter option, we set the range of publication as 2010–2017 for each author. The downloaded author-data was in comma separated-value format (CSV) which contained the list of all the papers belonging to an author along with the citations each of his/her papers received is shown in Figure 2 below.

The citation report option had further detail about the paper of the author in consideration being cited, the author of the citing paper, the university the author belonged to and the year of publication along with *h*-index and total publications. We prepared a list of 35 authors with various *h*-index ranges

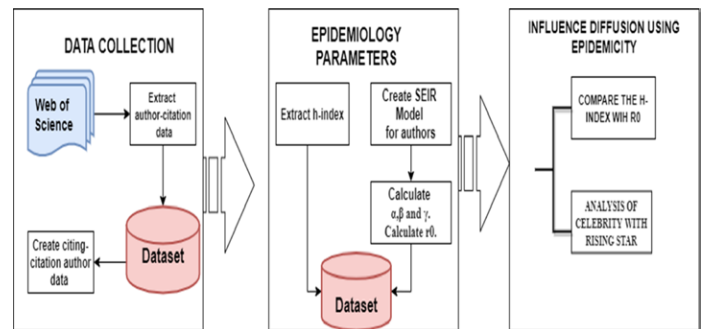


Figure 1: Data Flow and steps.

#	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE	AF	AG	AH	AI	AJ	
150														41	0	0	0	1	SPRINGER D
151														25	0	0	1	15	SPRINGER D
152														19	0	0	1	12	UNIV OSIE S
153														27	0	0	0	0	ELSEVIER S/A
154														34	0	0	0	0	SCIENCE & W
155														30	0	0	0	6	UNIV VIST T
156														21	0	0	1	9	SPRINGER N
157														12	0	0	2	9	KOREAN SCS
158														34	0	0	3	11	SPRINGER N
159														43	0	0	1	7	MORI AG B
160														15	0	0	0	2	IEEE INST EP
161														21	0	0	4	6	IEEE INST EP
162														14	0	0	0	12	SAGE PUBLI
163														35	0	0	4	6	IEEE INST EP
164														21	0	0	0	0	SAGE PUBLI
165														26	0	0	0	0	UNIV UTARS
166														12	0	0	0	1	IEEE INST IT
167														26	0	0	0	0	SCIENCE & W
168														42	0	0	4	11	IEEE INST EP
169														21	0	0	1	15	SCIENCE & D
170														8	0	0	0	0	ICC INTERK

Figure 2: Snapshot of citation CSV downloaded for R Buuya.

(less than 10, greater than 50) and downloaded data accordingly. The data collection steps are explained as follows:

1. In the first step we select an author 'i' who we are trying to analyze.
2. In the second step we are taking in the data of the articles citing the work of the author 'i' and call the set of all these articles as 'Citing Articles'.
3. In the third step we segregate the citations received year wise and compile the list of author-citation data year wise as shown in Table 1.
4. Next each citation is further investigated based on the source of the citation. We calculate how many papers of the author 'i' are going from one state to another, i.e, Susceptible to Exposed, Susceptible to Infected, Susceptible to Recovered, Exposed to Infected, Exposed to Recovered and Infected to Recovered. Each of these states has been described in next section and holds the same meaning.
5. In the final step we calculate the values of the various coefficients by taking the average of each of the transitions.

2. Transition from Original SEIR Model to Our Influence Diffusion Model

Scientometrics is often considered as the quantitative study of any measurable aspect of scientific activity (mainly those reflected in the scientific literature), with the aim of understanding and, if possible, improving its operating mechanism. There are two variations that are proposed in.^[20]

1. Structural scientometrics. Its purpose is the mapping of the structure of scientific communities, sets of documents, cognitive ideas, etc. Its typical techniques are, among others, graph theory, network analysis, cluster analysis.
2. Dynamic scientometrics. Its purpose is to describe the space-time behavior of scientific information through scientometric objects (authors, publications, citations, etc.). Its typical methodological tools are ordinary and partial differential equations, stochastic models, and computer simulation.

Epidemiological models in scientometrics have been used to describe the dynamic behavior of authors and citations.^[20,21] In our approach, we have concentrated on the SEIR model variant. This is because of the extended class 'Exposed' that suitably models our requirements of segregating citations based on the source. The flowchart of the SEIR model is given in Figure 3. The ovals in the flowchart represent the different compartments of the infection model namely Susceptible (S), Exposed (E), Infected (I) and Removed (R). The transition from one class of nodes to another is represented with the help of arrows and the rate of transitions are indicated.

Next the data collected as discussed in previous section is fed into our model as shown in Figure 4 with each state defined as follows: -

1. Susceptible (S): The susceptible class represents the set of published articles by an author in a particular year.

Table 1: Snapshot of Author-Citation Data year wise.

Author Name	2010	2011	2012	2013	2014	2015	2016	2017
Thouvenin	0	0	1	4	5	8	4	31
Sharifzadeh	0	0	0	0	5	22	62	76
Schildgen	36	37	43	55	52	78	80	126
Marchesan	26	31	26	53	66	90	144	202
Kim K	0	0	0	0	10	95	203	301
Hugelius	2	4	12	28	70	110	238	274
Garfinkel	14	18	35	48	63	126	169	228
Jgoodall	383	395	466	510	563	707	876	990
Stephen J	0	0	1	0	1	1	32	83
Timbernerslee	106	119	168	171	137	160	180	158
Millionfriedman	1546	1764	1778	2131	2226	2432	2861	3016
Dainfossey	7	1	6	7	7	7	15	5
Alfredkinsey	33	33	27	26	18	15	23	29
Abecasisgoncalo	2043	2638	3193	3489	3940	4296	4400	4591
Dennis	565	752	774	938	1115	1230	1265	1345
Buyya	97	114	145	240	319	627	810	1183
Akyildizif	638	756	809	998	1063	1245	1213	1102

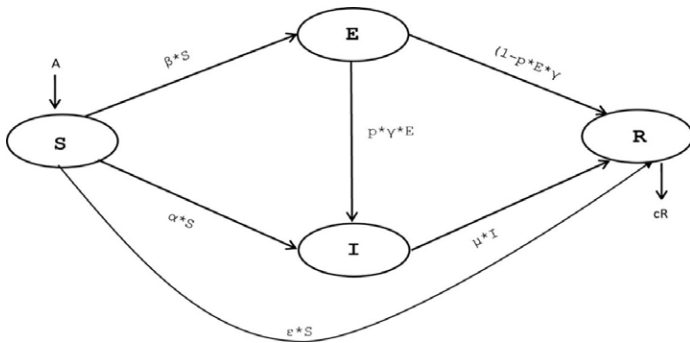


Figure 3: SEIR model with transition States.

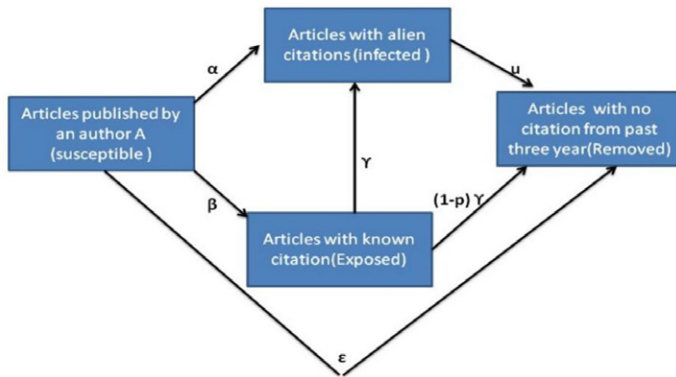


Figure 4: Transition from SEIR states to our Model.

2. Exposed (E): The exposed class (E) represents the set of articles which have only received local citations, which means citations from authors belong to the same institute/ organization or community as that of the published author.
3. Infectious (I): The infected class (I) represents the set of articles which have been cited by alien authors, meaning not belonging to the same institute or organization.
4. Removal (R): The recovered class (R) represents the set of articles which were not cited by any authors for a period of three years.

The size of the above classes w.r.t to the proportions of the network can be expressed as follows:

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

3. Vital Dynamics of SEIR Model: Parameters and Equations

Once the structural model was ready, next we start the process to describe the dynamic behavior of our model with the parameters redefined. The citations for the data collected is set as:- Initial condition (t=0) is set to year 2010 and time span is 8 years that is until 2017. The dynamism of SIER model for author citation can be represented using a set of nonlinear

differential equations based on the Initial Value Problem (IVP), in the following general form.

$$\frac{dY}{dt} = f(t, Y, \zeta), Y(t_0) = Y_0 \quad (2)$$

Where $Y \in IR^n, Y = Y(t)$ is a vector which defines the size of the composition of different classes of nodes. The size of the above classes' w.r.t to the proportions of the network can be expressed as in equation (2) where, N is the total number of the nodes in the network. Let $1/\lambda, 1/\gamma$ and $1/\mu$ be the average periods of time a node remains in Susceptible, Exposed and Infected classes, respectively. Assume λ to be infection rate at which the susceptible nodes acquire the infection. Assume that p is the probability for an exposed paper to become an infected paper (if cited by an alien author). Also, (1-p) is the probability that an exposed paper becomes re- covered (not cited for 3 a period of 3 years consecutively). We also assume that no new papers are written from the time the citations are being counted. A complete list of all the parameters used in the equations are listed below in Table 2.

1. The published articles of a researcher are in the susceptible state where any citation received from local network of an author indicates a transition of the articles to Exposed class with a rate of β . Similarly, another portion of the susceptible papers which are cited by alien authors and with a rate of ' α ' move to the infected class of papers. Certain papers may not be cited at all for a period of 3 years since the paper has been published. These papers with a rate of ϵ move to the recovered class. Figure 4 shows the complete transition of the above. So, the rate of change in size of susceptible papers can be represented in terms of following differential equation

(3) Using the principle of the law of mass action.

$$\frac{dS}{dt} = -\lambda S + A = -(\alpha + \beta + \epsilon)S + A \quad (3)$$

2. The susceptible portion of papers infected move to the class of exposed, they will remain in this class for a period of the duration $1/\gamma$. The next transition is either to infected state at

Table 2: Table of Parameters.

Symbol	Description	Unit
λ^{-1}	Average susceptible period	(year)
β	Exposure rate	(year) ⁻¹
ϵ	Average rate of at which paper travels from S to R	(year) ⁻¹
α	Average rate of at which paper travels from S to I	(year) ⁻¹
γ^{-1}	Average exposure period.	(year)
μ^{-1}	Average infectious period	(year)
A	Recruitment Rate	(year) ⁻¹
p	Probability of an exposed to become infected	Dimensionless

a rate proportional to γ with a probability p or some of them may not be cited for a period of 3 consecutive years and may go to the re-covered class at a rate proportional to $(1-p)/\gamma$. It follows that the equation expressing the rate of change in size of the exposed proportion of the population is:

$$\frac{dE}{dt} = \beta S - p\gamma E - (1-p)\gamma E \tag{4}$$

3. The average life of infected papers is taken as $1/\mu$, so the rate with which a paper goes from infected class to recovered class is given by μ . The equation for rate of change in size of infected paper is given by the average life of infected papers is taken as $1/\mu$, so the rate with which a paper goes from infected class to recovered class is given by μ . The equation for rate of change in size of infected paper is given by:

$$\frac{dI}{dt} = \alpha S + p\gamma E - \mu I \tag{5}$$

4. As we have described earlier, if a paper in any of the class has not been cited for a consecutive period of 3 years, it'll move to the recovered state. The 'cR' is the rate at which papers are removed from the model. The differential expression for Recovered state is given by:

$$\frac{dR}{dt} = \epsilon S + (1-p)\gamma E + \mu I - \alpha R \tag{6}$$

The probability of an exposed paper going to infected class has been assumed as a small value of 0.001. The recruitment rate is equal to the rate at which the paper leaves the recovered state R. We assume that this is periodic in nature.

$$A = \alpha R \tag{7}$$

These four ordinary differential equations coupled with the flowchart form a system that governs the dynamics of author citation. A solution to this system is a vector function that provides, at any time t , the coordinates of the point in four-dimensional space whose components are expressed in terms of sizes of susceptible, exposed, infected, and recovered respectively. From the system of ordinary differential equations, it follows that:

$$\frac{dS}{dt} + \frac{dE}{dt} + \frac{dI}{dt} + \frac{dR}{dt} = 0 \tag{8}$$

Let the sum of the 4 ordinary differential equations be given by S_0 : -

$$= -(\alpha + \beta + \epsilon)S + A + (\beta S - p\gamma E - (1-p)\gamma E) + (\alpha S + p\gamma E - \mu I) + (\epsilon S + (1-p)\gamma E + \mu I + \alpha R)$$

From equation 3,4,5 and 6 we get.

$$\begin{aligned} &= -\alpha S - \beta S - \epsilon S + A + \beta S - p\gamma E - \gamma E + p\gamma E + \alpha S + p\gamma E - \mu I + \epsilon S + \gamma E - p\gamma E + \mu I + \alpha R \\ &= -\alpha S - \beta S - \epsilon S + A + \beta S - p\gamma E - \gamma E + p\gamma E + \alpha S + p\gamma E - \mu I + \epsilon S + \gamma E - p\gamma E + \mu I + A \end{aligned}$$

= 0 (By cancelling the negative and positive terms)

Integrating Equation 8. with respect to the time, the integral yield the following result

$$S(t) + E(t) + I(t) + R(t) = k \tag{9}$$

Where k is the constant of integration which is equal to $1at$ the initial time $t = 0$. Epidemiologically, we can conclude that the size of the population under consideration does not change during the whole period of author citations.

$$S(t) + E(t) + I(t) + R(t) = 1 \tag{10}$$

All states of our model are now defined with equations and the model is balanced as well.

REPRODUCTION NUMBER: AN INDICATOR FOR INFLUENCE DIFFUSION

In the literature of infectious diseases, the basic reproduction number R_0 is the determinant parameter for the spread of the disease. The most important uses of R_0 are determining if an emerging infectious disease can spread in a population. This property is suitably used in our model to simulate and measure the spread of scholastic influence. As far as the threshold conditions of influence propagation is concerned, to spread out, the basic reproduction number must be greater than 1, otherwise influence propagation will die off, i.e., $R_0 > 1$. The derivation of R_0 for our model begins by identifying the various factors affecting the spread of influence. The spread of influence of an author depends on the average rate of citations received from alien sources, duration of the infection and transmittable rate. This is suitably modelled as: -

$$R_0 \propto \left(\frac{\text{infection}}{\text{contact}} \right) \cdot \left(\frac{\text{contact}}{\text{time}} \right) \cdot \left(\frac{\text{time}}{\text{infection}} \right) \tag{11}$$

More specifically:

$$R_0 = \tau \cdot \bar{c} \cdot d \tag{12}$$

Where τ is the transmissible (i.e., probability of infection given contact between a susceptible and infected individual), \bar{c} is the average rate of contact between susceptible and infected individuals, and d is the duration of infectiousness. We can

now define the variables on the RHS of the equation (12) in terms of the various parameters shown in Table 2 as follows: -

$$\tau = \frac{1}{(\alpha + \gamma)} \tag{13}$$

Here the probability of the infection given that contact happens is found by adding and then taking inverse of the rates of the two ways a paper can get infected that is by going from Susceptible to Infected State and Exposed to Infected State as shown in equation (13)

$$\bar{c} = \frac{\beta + \alpha + \gamma}{3} \tag{14}$$

Here we need to take the average of the rate of contact between susceptible and infected individuals which includes rate of Susceptible to Infected, Exposed to Infected as well as Susceptible to Exposed as shown in equation (14). We also consider Susceptible to Expose as getting exposed increases the papers chance of getting infected.

$$d = \frac{1}{(\epsilon + \mu)} \tag{15}$$

We now use the Equations 13,14 and 15, to derive Reproduction Number R0 as: -

$$\begin{aligned} &= \tau \cdot \bar{c} \cdot d \\ &= \left(\frac{1}{(\alpha + \gamma)} \right) \times (\bar{c} \cdot d) \\ &= \left(\frac{1}{(\alpha + \gamma)} \right) \times \left(\frac{\beta + \alpha + \gamma}{3} \right) \times (d) \\ &= \left(\frac{1}{(\alpha + \gamma)} \right) \times \left(\frac{\beta + \alpha + \gamma}{3} \right) \times \left(\frac{1}{(\epsilon + \mu)} \right) \\ &= \left(\frac{\beta + \alpha + \gamma}{(3 \times \alpha + \gamma) \times (\epsilon + \mu)} \right) \end{aligned}$$

Therefore, R0 for our model is given as:

$$R_0 = \left(\frac{\beta + \alpha + \gamma}{3 \times (\alpha + \gamma) \times (\epsilon + \mu)} \right) \tag{16}$$

VII. RESULTS OBTAINED AND VALIDATION

We construct our SEIR influence diffusion model for each of the 35 authors. The citations of each author are investigated and accordingly the states and transitions are made. The value of each coefficient α , β , γ , μ are derived using the Algorithm 1.

For example, as shown in Table 3, for an author with 24 papers we carefully analyze per paper/citations and year wise the movements.

Algorithm 1: Calculation of parameters.

- 1: Input: File containing all the citations to the author.
- 2: Output: The co-efficient of the model for the author
- 3: Select an Author i and get the authors citing articles.
- 4: Let citing Articles be the list of all articles citing the author.
- 5: for each citing article ca in citing Article considered do.
- 6: for each paper p cited by ca do.
- 7: for each year y in consideration do
- 8: Calculate the number of papers moving from Susceptible. state to Exposed state and add it to S to E[y]
- 9: Calculate the number of papers moving from Susceptible. state to Infected state and add it to S to I[y]
- 10: Calculate the number of papers moving from Susceptible. state to Recovered state and add it to S to R[y]
- 11: Calculate the number of papers moving from Exposed. state to Infected state and add it to E to I[y]
- 12: Calculate the number of papers moving from Exposed. state to Recovered state and add it to E to R[y]
- 13: Calculate the number of papers moving from Infected state to Recovered state and add it to I to R[y]
- 14: end for
- 15: end for
- 16: end for
- 17: for each [S to I; S to E; S to R; E to I; E to R; I to R] do
- 18: Sum all the values and divide by number of terms to get the coefficients

Table 3: Year wise movement of papers from S to I.

Year	No.of paper from S to I
2010	8
2011	0
2012	4
2013	0
2014	4
2015	4
2016	0
2017	4

The number of papers moving from susceptible state to infected year wise is as follows: - The value of α is obtained by dividing total papers moved from susceptible to infected state by the number of years. So, the value obtained is $24/3=3$. Similarly, all other coefficients are also derived and stored in the csv file as shown in Table 4.

The reproduction number is found to be unaffected by the number of papers which indicates that any researcher with higher publication count and h -index does not ensure that his/her influence will be high. The R0 values in Table 5 are consistent with the ideology of the model which is built to place authors at higher ranks, if their eminence is accomplished without the aid of external influence or collaboration of any kind. Certainly, Friedman M and Hawking J deserve to get the top 2 positions. The authors are very well recognized

Table 4: Snapshot of Complete Data with all Parameters.

No of Papers considered	Authornames	Beta	Alpha	Epsilon	Gamma	Mew	h-Index	R0 Values
233	Alkydiz K	0.625	28.5	0.001	0.5	0.125	66	2.7025
299	Buyya R	0.5	28.125	0.001	0.375	0.001	80	169.5906
125	Dennis J	0.25	15.375	0.001	0.25	0.001	100	169.3333
43	Goncalo A	0.125	5.25	0.001	0.125	0.001	108	170.54
26	Kinsey	0.001	3.25	0.001	0.001	0.001	101	166.66
21	Fossey D	0.5	2.125	0.001	0.25	0.001	83	160.54
1396	Friedman	66.125	108.375	0.001	34.75	0.001	94	243.668
139	Tim B	0.001	17.375	0.001	0.001	0.001	87	166.666
198	Hawking J	0.625	25.25	0.001	0.001	0.001	90	170.79
174	Goodall	0.5	21.25	0.001	0.001	0.001	61	170.588

Table 5: Analysis of R0 values for various celebrity authors.

No. of papers	Author name	h-index	R0
43	Goncalo A	108	170.54
26	Alfred K	101	166.66
125	Dennis J	100	169.33
1396	Friedman M	94	243.668
198	Hawking J	90	170.79
21	Fossey D	83	160.54
299	Buyya R	80	169.59
233	Alkydiz	66	2.705
174	Goodball	61	170.588

Table 6: Analysis of R0 values for various Rising authors.

No. of papers	Authorname	h-index	R0
5	Dane D	10	222.22
14	Allison K	9	179.487
28	Harrison Chris	9	166.667
20	Ordinary I	9	0.666
4	Sahar S	8	166.66
3	Olivier T	6	166.66

researchers in computer science domain. It is remarkably interesting to explore the R0 values attained by Dane D and Olivier T. They have reasonably higher R0 scores since their infection rates are higher than many authors with higher h-index. Also seen is author Buyya R the number of papers is 299 and R0 is 169 indicating his work is influential and epidemic in nature. On the other hand, Alkydiz K has high h-index and R0 > 1 indicating epidemic yet less influential. Table 6 highlights few authors whose R0 value is compared with h-index and number of papers published. It indeed is interesting to note that author Olivier T has just 3 publications but the R0 is high owing to high infection rate and low removal rate. These authors are suitably termed as "Rising star".

The SEIR curves are plotted for various authors with high h-index, low h-index and a comparative study is performed between Infection rate vs h-index.

The Figures 5 and 6 clearly indicate that the infection rate in both the authors are found high as many citations where received from outside community. Reproduction number is a clear indication of spread of infection and here it is

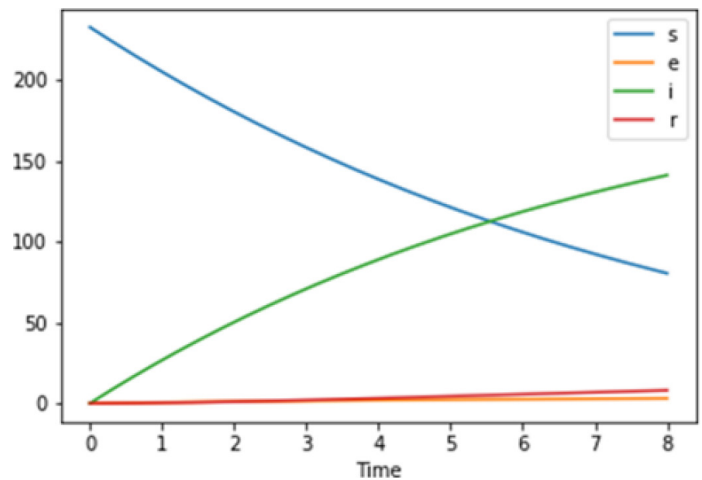


Figure 5: SEIR Curve for a Celebrity Author with h-index of 66.

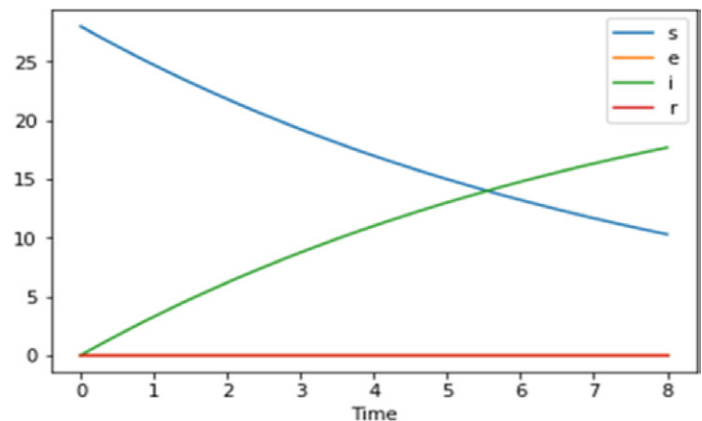


Figure 6: SEIR Curve for a Rising Author with h-index 8.

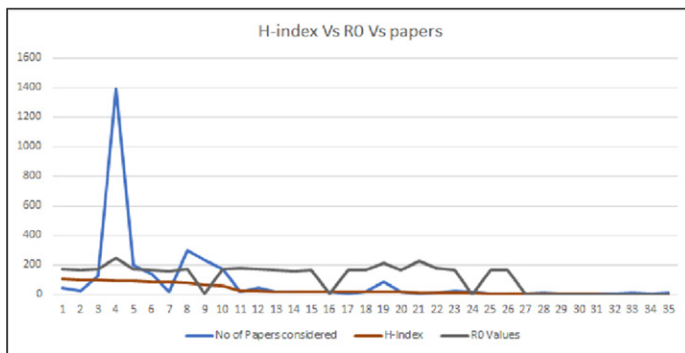


Figure 7: Paper considered vs R0 vs h -index.

independent of the h -index or the publication counts. The rate of infection is higher in these authors and they are called as super spreaders. With our analysis using SEIR model it is significantly found the even researchers with less publication rate can have high infection rate and become influential. Figure 7 clearly shows that there no relation between R0 and h -index or Number of papers published. Our model appropriately indicates that influence of an authors must be qualitatively measured.

CONCLUSION

The field of bibliometric has developed amazingly through the span of decades, implying that researchers are now dealing with more tougher questions. The assumption involved is that the citations received by a publication all have the same value when some of the citing publications would clearly have impacts different than others. The question becomes how such a convention could be set aside, making way for a new paradigm. Our effort is to clearly demarcate between citations within community and outside community. These citations are then suitable named exposed and infected states. The mathematical modelling of transmission of infection is suitably used to create a new model that can substantially work on "influentiality" as a parameter for observing an author's growth. The work focuses on quality of citation rather than the number of citations received. This ensures that all the researchers work is weighed equally without any biases and quality is the paramount winner. The contribution of this paper could lead to several consequences like

1. It could help identify suitable candidates for national and international research awards based on R0 scores.
2. It could help determine interdisciplinary Influence of a scholar.

3. Potential" Internationality" of an author could be measured.
4. Institutional growth in non-opportunistic research can be traced based on the fraction of its faculty boasting of high influence scores.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Schubert A. Scientometrics: The research field and its journal. In: Organizations and Strategies in Astronomy. Dordrecht: Springer; 2001; 179-195. DOI: 10.1007/978-94-010-0666-8-12.
2. Wang W, Street WN. Modeling and maximizing influence diffusion in social networks for viral marketing. Applied Network Science. 2018;3(1):1-26. DOI: 10.1007/s41109-018-0062-7.
3. Jain R. The Art of Computer Systems Performance Analysis. New Delhi: John Wiley & Sons; 2008.
4. Sobin CC, Saha S, Raychoudhury V, Fidele H, Sinha S. CISER: An Amoebiasis inspired Model for Epidemic Message Propagation in DTN. arXiv preprint, arXiv: 1608.07670. 2016
5. Buchandiran G. An Exploratory Study of Indian Science and Technology Publication Output, Department of Library and Information Science, Loyola Institute of Technology Chennai. 2011. <http://www.web-pages.uidaho.edu/mbolin/buchandiran.htm>
6. Garfield E. Citation indexes for science. Science. 1955;122(3159):108-11. doi:10.1126/science.122.3159.108
7. Kao C. The Authorship and Internationality of Industrial Engineering Journals. Scientometrics. 2009;80(3):123-36.
8. Neelam J, Snehanthu S, Siddhant G, Mukunda RJ. Ranking of Journals in Science and Technology Domain: A Novel and Computationally Lightweight Approach. IERI Procedia. 2014;10:57-62 doi: 10.1016/j.ieri.2014.09.091
9. Ginde G, Saha S, Mathur A, Venkatagiri S, Vadakkepat S, Narasimhamurthy A, Sagar BD. ScientoBASE: A framework and model for computing scholastic indicators of non-local influence of journals via native data acquisition algorithms. Scientometrics. 2016;108(3):1479-529. DOI: 10.1007/s11192-016-2006-2.
10. Chang CL, McAleer M, Oxley L. Coercive journal self citations, impact factor, journal influence and article influence. Mathematics and Computers in Simulation. 2013;93:190-7.
11. Garfield E. Citation analysis as a tool in journal evaluation. Science. 1972;178(4060):471-9.
12. Svensson G. Scholarly journal ranking (s) in marketing: single- or multi-item measures?. Marketing Intelligence & Planning. 2008;26(4):340-52.
13. Buela-Casal G, Perakakis P, Taylor M, Checa P. Measuring internationality: Reflections and perspectives on academic journals. Scientometrics. 2006;67(1):45-65.
14. Franceschet M, Colavizza G. Timerank: A dynamic approach to rate scholars using citations. Journal of Informetrics. 2017;11(4):1128-41.
15. Woo J, Chen H. Epidemic model for information diffusion in web forums: experiments in marketing exchange and political dialog. SpringerPlus. 2016;5(1):1-9. DOI: 10.1186/s40064-016-1675-x
16. Al-Sheikh SA. Modeling and analysis of an SEIR epidemic model with a limited resource for treatment. Global Journal of Science Frontier Research Mathematics and Decision Sciences. 2012;12(14):56-66.
17. Langville AN, Meyer CD. Google's PageRank and beyond: The science of search engine rankings. USA: Princeton University Press; 2011.
18. Haddow G, Genoni P. Citation analysis and peer ranking of Australian social science journals. Scientometrics. 2010;85(2):471-87.
19. Moed HF. Measuring contextual citation impact of scientific journals. Journal of Informetrics. 2010;4(3):265-77.
20. DST India. Bibliometric study of India's Scientific Publication outputs during 2001-10: Evidence for changing trends, Department of Science and Technology, Government of India, New Delhi; 2012.
21. Das AK, Mishra S. Genesis of Altmetrics or Article-Level Metrics for Measuring Efficacy of Scholarly Communications: Current Perspectives. Journal of Scientometric Research. 2014;3(2):82-92.