Journal of Scientometric Res. 2022; 11(1):79-88
http://www.jscires.org

Research Article

NLRIS: Modeling and Analysis of Non-local Influence of Research Output of Institutions

Poulami Sarkar[1], Gambhire Swati Sampatrao[2,*], Snehanshu Saha[3]

[1]Department of Informatics, TU Munich, GERMANY.
[2]Department of CSE, PES University, Bengaluru, Karnataka, INDIA.
[3]Computer Science and Information Systems and APPCAIR, BITS Pilani KK Birla Goa Campus, Goa, INDIA.

ABSTRACT
In this paper, we posit a novel metric for ranking academic institutions based on the extent of influence their publications have on the global research community. We quantify the university's global research impact using a calculated internationality score we call 'Non-local Research Influence Score' (NLRIS). We propose the evaluation metric which is fair to the institutions which are smaller in size and do not qualify to be considered for the reputed ranking schemes but are performing well in terms of research quality and attracting citations and collaborators around the world. Our work calculates the scores of 75 universities using the Constant Elasticity of Substitution (CES) function. This score is calculated as a function of the global influence in terms of citations (Non-local Influence Quotient - NLIQ) and the extent of international research collaboration International Collaboration Ratio - ICR). The CES function accepts 2 variables, K and L as input and produces the variable Q as an output. In our model, we substitute K and L with the calculated values of ICR and NLIQ. The output of the CES function provides us with the desired internationality score of each institution. PSO is used to model the constrained optimization problem. In this paper, we also define a metric for representing NLIQ and ICR of individual universities. We have also performed a comparison between QS ranking and our calculated internationality scores.

Keywords: Constant Elasticity of Substitution, Non-local influence quotient, International collaboration ratio, Concavity, Internationality, Influence.

Correspondence
Gambhire Swati Sampatrao
Department of CSE, PES University, Bengaluru-560085, Karnataka, INDIA.
Email id: jagdalesp@yahoo.com

Received: 18-08-2021;
Revised: 09-02-2022;
Accepted: 19-03-2022.
DOI: 10.5530/jscires.11.1.8

INTRODUCTION

In today's world, academic research is one of the most dynamic areas with hundreds of papers being published by almost every established university every year; influencing not just the research community but also the quality of the institute on the whole. As a consequence, any research that aims to qualitatively evaluate the quality of academic research is a non-trivial task. Every year, open-source services such as AMiner, Google Scholar, Web of Science, etc compile massive amounts of scientometric data from such research publications. For our research, we have collected university-wide publications data and ranked the universities based on the impact their publications have on the global research community.

In our paper, we adopt a meta-heuristic model-based approach for ranking universities based on their extent of international collaboration and the global impact of their research work. This kind of work can be of great interest to many potential students, academicians as well as researchers. Most of the currently existing ranking metrics such as QS ranking, consider a vast multitude of factors for ranking universities, that may not always be relevant to a potential researcher. Moreover, they are sometimes heavily biased towards larger institutes that have a high faculty count and offer a wide range of sub-courses in a particular stream. In the light of this, our work offers a solution that ranks universities purely based on the impact of their research output and the scope of international collaboration.

Motivation and Problem Statement

The main intent of this manuscript is to analyze and quantify the influence-spread of the institution in the global spectrum; outside its peer group and country/continent. In recent years, there has been a major catalyst to measure the quality of research at all levels, from the individual paper right up to the institution. Quantifying and measuring the influence of the institution's research output is of great importance. These measures could be used by government agencies, funding bodies, research scholars, and institutions themselves to determine whether it is producing "enough" of the influence. Yet the attempt to measure the influence spread of the institution's research is sparse. Quantification of research output is an important parameter in ranking institutions. The institution's performance evaluation should be majorly based on institutional research performance. Existing institution ranking schemes consider parameters such as management of technology and infrastructure, faculty-student ratio, perception, pedagogy, etc., and hence do not truly reflect the non-local research influence spread of the institution.

We are specifically motivated to investigate the research quality of small institutions (in size) and the reasons for those institutions not featuring in reputed ranking sites. Consequently, we attempt to recognize those institutions based on research quality and attraction quotient (ICR) even though, it might be difficult to 'market' them for a variety of reasons, political included.

This motivates us -

- To investigate and select very specific research metrics that can reflect the extent of influence spread outside a university's peer or local group.
- To measure the amount of collaboration across countries/ continents for such an evaluation.
- To check and analyze if there are any anomalies between our computed scores and existing ranking schemes.

In this manuscript, we propose two research-specific metrics that account for citations received by the institution from outside its home country and the amount of collaboration across the countries. We compute the 'Non-Local Research Influence Score' (NLRIS) and with further post-facto analysis investigate the standard public perception that the institutions big in size perform well at the research front as well and gain the non-local influence. It is a known fact that significant publications have a far-ranging impact on research contributions in the subsequent years. Hence, we have defined 2 metrics that represent this impact and have used the same to compute the NLRIS of institutions. We call these 2 metrics:

1. ICR - International Collaboration Ratio

We have defined this metric based on the assumption that universities that have a high ratio of collaboration with a large number of international universities outside their own country tend to attract attention in the global research community. The percentage of ICR is an expression of interest among the peers to collaborate with a particular institution.

2. NLIQ – Non-Local Influence quotient

This metric is a ratio of the number of citations a paper receives from outside of its own country to the total number of citations. The NLIQ of a university is a measure of the amount of global attention that the university's research output receives.

Contribution Summary

In this paper, we propound[1] the following research contributions:

1. Two new metrics of scholarly impact and reach, along with the algorithms for computing them. These metrics are the non-local influence quotient (NLIQ) and international collaboration ratio (ICR).
2. A novel model that uses the two metrics, NLIQ and ICR as input and computes a score for ranking institutions. This model awards a Non-Local Research Influence Score (NLRIS), to each institute solely based on the global impact of their research and extent of global collaboration. This score is irrespective of the size of the institutions and is hence unbiased towards larger institutes with higher faculty count.
3. A comparison of our ranking scheme with that of other popularly used schemes such as QS ranking.
4. A comparison of the mismatch between the ranking scheme we proposed and that of QS ranking.

LITERATURE REVIEW

The presumption that publishing is the only indicator of research productivity has shifted in recent years to recognize the need for quantitative measures of research impact. Various bibliometric metrics are increasingly being used to assess an institution's research output. Researchers have tried several approaches to evaluating the institution's research output.

Aithal PS and Kumar PM [1] argued that the Institutional Research Performance should be used to evaluate institutions' performance. The authors proposed a model focused on institutional variables such as the number of papers, books, book chapters that are written, case studies, and the number of full-time faculty to calculate research productivity using an institutional research index. Madhan M, *et al.*[2] investigated that the impact factor and *h*-index syndrome affects government departments, funding bodies, academic, and research institutions, and as a result, research evaluation in India has trouble separating good science from poor. Authors suggested that when recruiting new professors, their research ideas and contributions to the field, as well as their originality and ingenuity, should be considered in addition to citation and publication counts. In their work, Docampo *et al.*[3] proposed scale-free and size-dependent steps to assess countries' and institutions' scientific contributions. They also keep track of gaps in performance across various research domains, recognizing institutions' and countries' strengths and weaknesses. This paper introduces a qualitative metric

1 https://github.com/Poulami-Sarkar/NLRIS-Modeling-and-Analysis-of-Non-localInfluence-of-Research-Output-of-Institutions

called Global Research Output (GRO), which is defined as the number of citations per publication divided by the number of publications. Massucci FA and Docampo[4] have used Page Rank to assess academic credibility through a citation network. The author assumes that citations represent a reference's credibility, so the Page Rank algorithm can be used to determine rankings. A similar Page Rank-based approach is explored by Nykl M *et al.*[5] to rank authors based on citation network analysis. Kazi P, Patwardhan M and Joglekar P[6] proposed context-based quality metrics. The authors analysed the text surrounding the self-citations and citations for semantic analysis. The semantic similarity was also determined. Similarly, West JD, *et al.*[7] presented the method by which the journal Eigenfactor is used to rate institutions in their work. The citation network's most significant nodes are indicated by the Eigenfactor ranking. The same methodology takes into account eigenvector centrality and computes Eigenfactor scores for authors using an algorithm close to page rank. The sum of the authors' scores can be used to measure the institution's ranking. Abramo G,[8] explained the significance of the term "impact" The contribution of research output to further scientific and technological advancement, defines impact. Measuring internationality is addressed by Ginde G *et al.*[9] In their paper, authors have proposed a methodology that measures 'internationality' by removing local effects to define influence. The metric Non-Local Influence Quotient (NLIQ) is presented and Cobb–Douglas Production Function is utilized.

Dataset

We sourced our data sets from Web of Science. We selected 75 institutions from The Times Higher Education Rankings and collected the metadata of articles published by these universities in the domain of Computer Science in the year 2012-13. We have also created a domain-specific dataset of 50 institutions for the 'Robotics' domain. To compute model input parameters NLIQ and ICR, we extracted the source article's country, citations received from other countries, authors, and their affiliations from the Web of Science. Download the data in the plain text form for further processing to compute ICR and NLIQ.

1. NLIQ

Defining and measuring 'Internationality' at the institution level should be based on its non-local impact. In large institutes, that have many faculties, it is often common to find tight-knit research communities that mutually influence each other's work. Since most common metrics that are used for measuring impact, consider the overall impact, they do not take into account that a significant portion of the impact may be on account of these local groups. As a direct consequence, any ranking scheme that uses such a metric to award a rank will undoubtedly be biased towards certain institutes. In an attempt to mitigate this issue, we propose a metric known as Non-Location Influence Quotient or NLIQ, that quantifies the research impact of the institute solely based on their global citations.

$$L = \frac{\sum_{p \in P(a)} \text{No\_of\_global\_citation\_for\_paper\_p}}{\text{Total\_citations\_on\_paper\_p}} \quad (1)$$

2. ICR

The international Collaboration Ratio (ICR) is a novel metric that quantifies the extent to which each university collaborates with international collaborators. For each university, ICR can be broadly defined as the ratio between the weighted sum of the contributions of domestic authors and the weighted sum of the contributions of international collaborators.

A high value of ICR would imply that researchers from across the globe are interested in collaborating with a particular university. This is undoubtedly a tacit acknowledgment that the research output produced by an institute is of high value to the research community. ICR calculation requires the following fields for each paper for each university.

- Number of authors who have authored the paper
- A number of authors belonging to different countries are grouped by country. Eg. 2 authors from India and 4 authors from the USA.
- (Optional) The number of authors belonging to the same countries but different universities grouped by the university.

ICR is computed as follows:

1. Authors ratio, $ar_i \leftarrow$ count(countries)/count(authors(r_i))

2. $w_i \leftarrow ar_i / \text{count}(\text{authors}(r_i) \in u_i)$

The above formula calculates the weight of the article, u_i indicates the university to which the authors of the article belong.

3. $nw_i \leftarrow w_i / \text{authors}(r_i)$

4. $icr_i \leftarrow \Sigma nw_i$ / count(r_i with international collaboration)

where r_i is university record, nw_i is normalized weight and w_i is weight.

Our Model

We propose a solution for quantifying the global influence of an academic institute by assigning them a score that is calculated using the ICR and NLIQ values, discussed in the previous section. We call this score the Non-Local Research Influence Score (NLRIS) of an institution. The NLRIS score signifies the impact an institute's publications have on the global research community.

In our work, we use the Constant Elasticity of Substitution or CES function to calculate the internationality score. This function accepts 2 variables, K and L as input and produces the variable Q as an output. In our model, we substitute K and L with the calculated values of ICR and NLIQ. The output of the CES function provides us with the desired internationality score of each institution. The CES function is defined as follows-

$$Q(L,K) = \gamma(\alpha K^{\rho} + (1-\alpha)L^{\rho})^{\eta/\rho}, \quad (2)$$

where

Q= Quantity of output (Internationality or NLRIS score)

K = ICR (International Collaboration Ratio)

L = NLIQ (Non- Local Influence Quotient)

ρ = Elasticity of substitution, γ is an endogenous parameter

α = Share parameter

The derivative of Q given by Q' can be written as:

By chain rule,

$$Q' = \frac{\eta Q}{R\rho\gamma}\left[\log\left(\frac{K^{\alpha k^{\rho}} L^{L^{\rho}(1-\alpha)}}{R^{R/\rho}}\right)\right] \quad (3)$$

By Taylor series expansion,

$$Q' = Q^{-\frac{\rho}{\eta}} \frac{\ln K^{\alpha} L^{1-\alpha}}{\ln Q} \quad (4)$$

CES derivation using Taylor series approximation

The general form of the Constant Elasticity of Substitution (CES) production function (CESArrow) for two inputs is

$$Q(L,K) = \gamma(\alpha K^{\rho} + (1-\alpha)L^{\rho})^{\eta/\rho}, \quad (5)$$

where Q= quantity of output (Internationality or NLRIS score), and represent input parameters. Define $\rho = \frac{s-1}{s}; s = \frac{1}{1-\rho}$, $\rho > 0$. Internationality[2] has as its limits the Cobb-Douglas production function, i.e.,

$$\lim_{\rho\to\infty} Q = \gamma K^{-\alpha} L^{\alpha-1} \quad (6)$$

Proof: We can rewrite the above equation as

$$Q = \gamma(\alpha L^{\rho} + (1-\alpha)K^{\rho})^{\eta/\rho}, \quad (7)$$

$$\frac{1}{\gamma} Q = (\alpha K^{\rho} + (1-\alpha)L^{\rho})^{\eta/\rho}$$

$$\frac{1}{\gamma} Q = \exp(\eta/\rho \cdot \ln[\alpha K^{\rho} + (1-\alpha)L^{\rho}]) \quad (8)$$

We consider first-order Taylor expansion centered at zero of the term inside the logarithm,

$$1 + \alpha\cdot\rho\cdot\ln(K) + (1-\alpha)\cdot\rho\cdot\ln(L) + O(\rho^2)$$

$$\alpha K^{\rho} + (1-\alpha)L^{\rho} = 1 + \rho[\ln(K^{\alpha}\cdot L^{(1-\alpha)})] + O(\rho^2). \quad (9)$$

Now, combining equations 8 & 9, we obtain

$$\frac{1}{\gamma} Q = [1 + \rho(\ln(K^{\alpha}\cdot L^{1-\alpha})) + O(\rho^2)]^{\eta/\rho} \quad (10)$$

Define ; $\tau = \frac{1}{\rho}; \rho \to 0; \tau \to \infty$. Therefore,

$$\lim_{\rho\to 0} \frac{Q}{\gamma} = \exp(\ln(K^{\alpha}\cdot L^{1-\alpha}))^{\eta}. \quad (11)$$

Consequently, we can write:

$$\lim_{\rho\to 0} Q = \gamma(K^{\alpha}\cdot L^{1-\alpha})^{\eta} \quad (12)$$

Assuming elasticity of scale , and constant of elasticity , we get

$$\lim_{\rho\to 0} Q = K^{\alpha}\cdot L^{1-\alpha}. \quad (13)$$

This is the Cobb-Douglas production function that can be used for non-zero input parameters.

$$Q^{\frac{\rho}{\eta}} = 1 + \rho[\ln(K^{\alpha} L^{1-\alpha})] \quad (14)$$

$$\frac{dQ}{d\rho} = Q^{-\frac{\rho}{\eta}} \frac{\ln K^{\alpha} L^{1-\alpha}}{\ln Q} \quad (15)$$

In the CES function, the value of the shared parameter, dictates the weight that each of the two parameters contributes toward the internationality score. However, as long as the shared parameters add up to 1, the problem remains concave, thus guaranteeing global optima. In our model, we fix the value of α (shared parameter) as 0.5 to give equal weight to both NLIQ and ICR. However, as long as the share parameters (α, $1-\alpha$) add up to 1, the model is guaranteed to return global optima. For example, if we choose α = 0.1, there won't be a major change in the final NLRIS values as the optima are

2 Internationality Score and NLRIS are to be interpreted as equivalent, in the context of the problem.

always found wrt the control parameters. The major challenge for our research was determining an optimal value of ρ that would allow us to complete our calculation of NLRIS. We have computed this value using the constrained particle swarm optimization algorithm, the motivation for which will be discussed in section 'Choice of the optimization method'.

Choice of the model

It is well-known that CES belongs to the family of neoclassical production functions.[9,10] The CES production function for two inputs can be represented in the form of Eqn. (2). Consider an enterprise that has to choose its input bundle (A, B, C, D) where A, B, C, D are the determining/control variables. In Economics, we determine the (global) cost of minimizing and profit-maximizing input bundles for such a production outlay. The enterprise wants to maximize its production, subject to cost constraints. We conceive the Non-local Research Influence as a production function where the revenue = internationality. Minimizing the cost and maximizing profit is thus equivalent to maximization of the Non-local Research Influence (i.e., production) subject to an appropriate choice of parameters including ρ. Our motivation for using the CES function stems from two fundamental properties of the CES function. Firstly, the CES model is concave and hence converges to a global optimum. Secondly, the CES function being additive in nature overcomes the problem of the function tending to zero when faced with any input parameters that are either very small or are missing. This type of problem is frequent in most multiplicative models such as the Cobb-Douglas production model. Additionally, the CES function also incorporates a special feature known as "elasticity of substitution", represented by the variableρ. ρ is typically a small value that aims to model small changes in the input conditions to reflect the latest value of internationality, which is subject to change in every subsequent year.

$$\rho = \frac{\sigma - 1}{\sigma}$$

$$\sigma = \frac{1}{1 - \rho}$$

We consider the following cases:

1. $\sigma = 0, \rho \to 0$

2. $\sigma = 1, \rho = 0$

3. $0 < \rho < 1$

$0 < \rho < 1$ makes the two inputs, ICR and NLIQ as perfect substitutes. Either one becoming zero makes the model depend solely on the other input. When the degree of substitution of, $\sigma > 1$, the model output becomes smoother with dependence on both input variables (ICR and NLIQ) and when $\sigma > 1$. $\rho < 1$ the suitable range for the model is: $0 < \rho < 1$ and therefore, this is the most suitable range for the optimization model with the elasticity ρ. We'll see later that this choice is further refined by additional computational consideration.

Choice of the optimization method

Our motivation for choosing a PSO meta-heuristic is driven by the problem of curvature violation that multi-variate functions are known to suffer from. Curvature violation is equivalent to the premature change in the shape of the function, even before the optima is reached. Therefore, for the functional forms considered in the general additive CES model, the difficulty of computing the analytical maximum (NLRIS) is due to 'curvature violations'. Since the model relies on theoretical guarantees of global optima, and the desired optima due to the curvature violation of the functional form are difficult to achieve, we rely on approximate methods of the first order. Curvature violation is a major issue in cases of flexible functional form, due to elasticity and share parameters, and is expected to be consistent with theory when estimations of input parameters and production function (Y, in this case) are required from a functional form. Along with that, the task of maintaining the flexibility of functional form is also necessary. The phenomenon sometimes arises due to the added local restrictions, or constraints, in the optimization problem. Since the internationality score i.e., NLRIS is obtained as a solution to a constrained (concave) optimization problem, we expect curvature violation due to the general practice of assuming smooth gradients along the functional form. So, if the curve changes sign abruptly, the gradient ascent, which is usually applied to find optima, would fail to detect the violation and report incorrect maximal point of ascent in the curve. This complexity is handled by computing global maxima theoretically and algorithmically for each university, exploiting intrinsic concavity of the functional form, and ensuring 'no curvature violation'. This is explicitly done by the iterative, metaheuristic method (replacing gradient ascent/descent method) described next.

Particle Swarm Optimization (PSO)[11] is a meta-heuristic algorithm for finding the global minima of a function. As the name suggests, the PSO algorithm draws its inspiration from the swarm behavior of a flock of fish or birds. It was traditionally designed for unconstrained inputs. It works by iteratively converging a population of randomly initialized solutions, called particles, toward a globally optimal solution. Each particle in the population has two properties; position and velocity. Throughout the optimization, every particle keeps track of its current position and the best solution it has encountered, called p^{best}. The particles use their associated velocities to traverse the search space. PSO is used to emulate another classical optimization method, called the gradient descent/ascent (GD). GD is also an iterative method to find the optima of multi-variate functions (identical to the type of function we're optimizing in this paper), continuous and differentiable everywhere. For a point x, we compute the updated position in the attempt to search for the local minima (against the direction of the gradient). Let us start with a guess x_0 for a local minimum of the function f, and consider the sequence $x_0, x_1 \ldots x_n$ such that the update rule is written as: $x_{n+1} = x_n - \eta \nabla f(x_n), n \geq 0,$ and the sequence of iterates converge to the local minima. is a heuristic, known as the learning rate, and chosen in such a way that the rate of learning is not too fast or slow. This is emulated by PSO for functions that might encounter rugged landscapes during descent or are prone to curvature violations. The descent learning explores the search space in a continuous manner unless it encounters a bottleneck. The minima we seek to discover is the (global) best solution that PSO aims to find as well. Thus, it is natural for the swarm to track the overall best solution, called g^{best}. Each iteration of the swarm updates the velocity of the particle towards its p^{best} and the g^{best} values. Let f(x) be the function to be minimized, where x is a d-dimensional vector. f(x) is also called the fitness function. Our focus in this work is restricted to adapting PSO for unconstrained optimization problems to constrained ones as well as mitigating the curvature violation and the complexity of handling multi-variate optimization problems. PSO handles this by eliminating the need to compute gradients explicitly. This also helps the method get rid of the assumption that the function in question is differentiable everywhere. Thus, PSO can be adapted to find approximate minima/maxima of functions with singularities and multiple local minima/maxima.

The primary objective of PSO was to solve optimization problems with single/several objectives (*devoid of exact, analytical optima*). PSO navigates the search space by embracing a *swarm* that is nothing but a population of particles. The swarm, guided by characteristic equations, attempt to converge to an optimum. The movement of the particles in the search space to discover the optimal solution is governed by the velocity and the position update equations. These are represented in the following manner:

$$v_i^{(t+1)} = \phi v_i^{(t)} + c_1 r_1 (p_i^{best} - x_i^{(t)}) + c_2 r_2 (g^{best} - x_i^{(t)})$$
$$x_i^{(t+1)} = x_i^t + v_i^{(t+1)}$$

where $\phi, c_1, c_2 \geq 0$. Here, x_i^t suggests the position of particle i at time t, v_i^t represents the velocity of particle i at time t, p_i^{best} is the best position particle i has attained, and g_i^{best} is the best position that the swarm has ever attained. Additionally, parameters related to cognitive learning and social learning need to be defined, which regulate the position and velocity of the particles and they are c_1, c_2, respectively; r_1, r_2 are random values sampled from U(0, 1) which contributes to the stochastic nature of the search process.

Thus, computing NLRIS of institutions is equivalent to solving a constrained optimization problem driven by the following mathematical structure:

$$Q = \gamma(\alpha \cdot L^{\rho} + (1-\alpha) \cdot K^{\rho})^{\frac{\eta}{\rho}}$$

The value of ρ is chosen to lie within $0 < \rho \leq 1$. The coefficients α, 1 – α must lie in and sum to 1, by construction to satisfy concavity. The value of η is constrained by the scale of production used, 0 < η < 1 under Decreasing Returns to Scale (DRS) and η = 1 under Constant Returns to Scale (CRS). CRS implies the rate of production is proportional to input and we have chosen η = 1 for our calculations. γ = 1 is set as a convenience.

Choice of the model parameter, ρ

We used the concept of fixed point to determine the range of values within which an optimum value of ρ can be guaranteed to be found. In mathematics, a fixed point of a function can be defined as the point in the function's domain where the input is mapped to itself by the function. It can be mathematically, represented as follows:

A point c is a fixed point of the function f if f(c) = c

Any function for which the fixed point can be found in polynomial time will be guaranteed to converge when optimized using any meta-heuristic algorithm. The derivative of the CES function, is a clear indication that the CES function is concave in nature and hence will converge to a fixed point. To accurately calculate ρ using a meta-heuristic algorithm like constrained PSO, we need to find some acceptable range of ρ within which the value of ρ can be restricted. For this, we use the Runge-Kutta method of numerical approximation. Runge-Kutta is a well-known method of numerical analysis that iteratively computes the fixed point of a function. By solving the CES using the first-order Runge-Kutta method we can show the function indeed converges to a fixed point.

1st order ODE

$$k_1 = f(x_n)\Delta t$$
$$k_2 = f(x_n + 0.5k_1)\Delta t$$
$$k_3 = f(x_n + 0.5k_2)\Delta t$$
$$k_4 = f(x_n + k_3)\Delta t$$
$$x_{n_1} = x_n + \frac{1}{6}(k_1 + 2k_2 + 2k_3 + k_4)$$

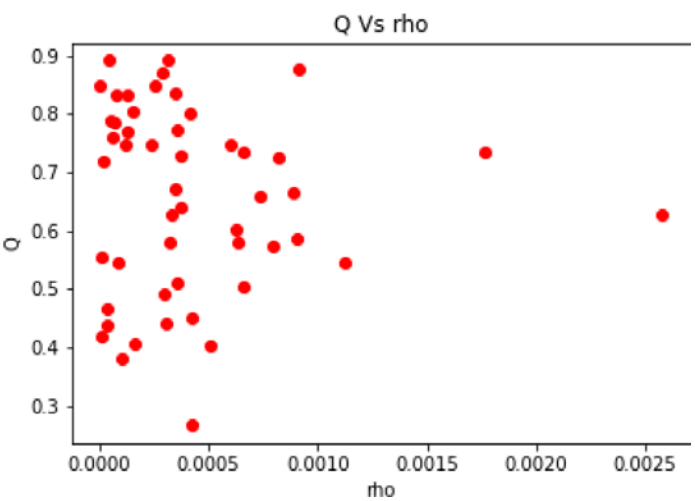

Figure 1: Graph of internationality, represented by Q VS ρ. This graph shows the distribution of internationality, Q with change in ρ.

Here, we assume f(x) to be

$$f(x) = \frac{dQ}{d\rho} = Q'$$

Figure 1 shows the plot of internationality Q with change in elasticity of substitution.

Figure 2 shows a trace of the cost history during PSO optimization. It is the learning curve when K (ICR) = 0.01 and L (NLIQ) = 0.6667. The values of ρ and internationality (cost) at the global optima is: ρ = 0.4382 and internationality (NLRIS) = 0.0006.

Figure 3 shows 3D learning curve for PSO optimization.

Result for constrained PSO optimization

The fixed point of the Runge-Kutta method to solve the Differential Equation governing our model yields the following ranges of values of ρ:

1. 0.0-1.1
2. 0.001-0.1
3. 0.1-0.9
4. 0.9-0.99
5. > 1

We have used the Runge-Kutta method of numerical approximation to estimate the fixed point of the CES function. We argue that the fixed point of the CES function which indicates acceptable ρ values for computing approximate optima of the model leading to the optimal NLRIS of universities should be at the intersection of the above range of values and the analytical range specified by the appropriate degree of substitution of the production model, from the Economics standpoint. Therefore, the choice of ρ is naturally in the range of 0.101 – 0.9 (i.e the CES function converges to a fixed point when the value of ρ is restricted between 0.101 and 0.9). Hence, we have chosen the parameters α = 0 OR α = 0.1; $\rho \in [0.1 - 0.9]$ for the optimization problem to be solved using constrained PSO.

Additionally, we also experimentally estimate the value of NLRIS by restricting the value of ρ between each of the five ranges listed above. The result of our experiment showed that the NLRIS values were consistent for most universities till rho < 0.9, proving that the model works in harmony for all ranges of ρ between 0 and 1. Additionally, our ranking scheme automatically awards higher ranks to universities having large values of NLIQ and ICR. This result makes it evident that our model gives more weight to linearly independent parameters, and consequently awards higher ranks to universities with large values of NLIQ and ICR. This can be observed in the case of the Hanoi University of Science and Technology where the values of both NLIQ and ICR are significantly high.

Table 1: This table shows the NLRIS values for 10 institutes where 0.101 < ρ < 0.9. The value of alpha=0.1.

Institute	ICR	NLIQ	NLRIS	Rank
Hanoi University of Sc. and Tech	0.68	0.95	0.92	2
University of Gothenburg	0.62	0.90	0.87	6
University of Zurich	0.57	0.87	.83	11
University of Helsinki	0.56	0.87	0.84	10
University of Vienna	0.58	0.83	0.80	20
University of Birmingham	0.57	0.82	0.79	22
Humboldt University of Berlin	0.59	0.76	0.74	38
Uppsala University	0.48	0.93	0.87	5
Trinity College Dublin	0.50	0.89	0.84	9
Brown University	0.49	0.57	0.56	66

Table 2: This table shows the NLRIS values for 10 institutes where 0.101 < ρ < 0.9. The value of alpha=0.5.

Institute	ICR	NLIQ	NLRIS	Rank
Hanoi University of Sc. and Tech	0.68	0.95	0.81	2
University of Gothenburg	0.62	0.90	0.75	4
University of Zurich	0.57	0.87	0.71	9
University of Helsinki	0.56	0.87	0.70	11
University of Vienna	0.58	0.83	0.70	12
University of Birmingham	0.57	0.82	0.69	13
Humboldt University of Berlin	0.59	0.76	0.67	17
Uppsala University	0.48	0.93	0.67	16
Trinity College Dublin	0.50	0.89	0.67	18
Brown University	0.49	0.57	0.53	59

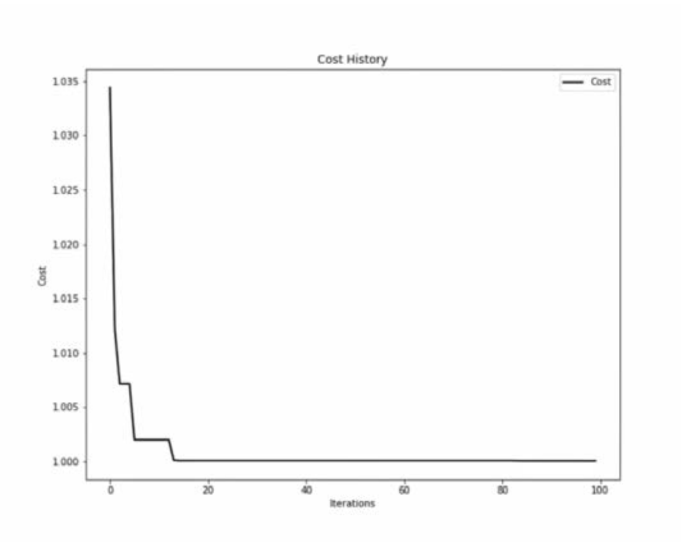

Figure 2: 2D Learning curve for PSO optimization. This graph shows how converges to its value at the fixed point in a 2D space.

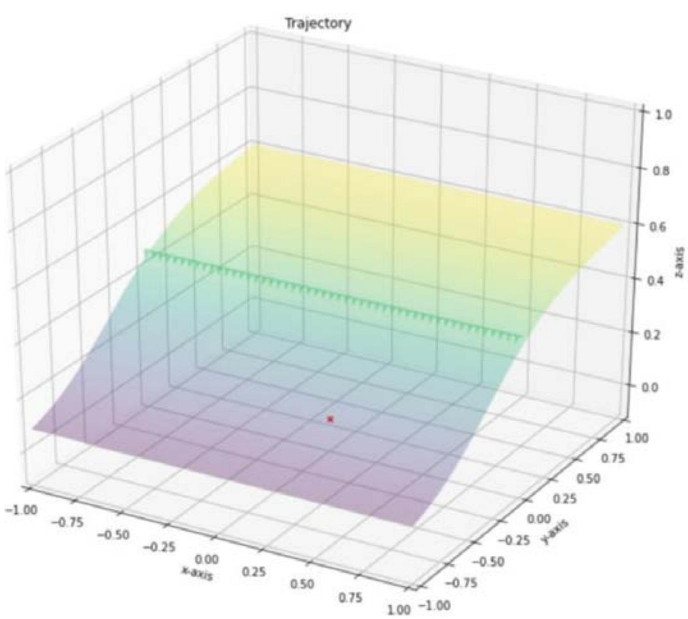

Figure 3: 3D Learning curve for PSO optimization. This graph shows how converges to its value at the fixed point in a 3D space.

Table 1 and 2 show the NLIRS scores for 10 institutes with the value of α=0.1 and α=0.5. From Tables 1 and 2 even with a significant change in the value of α, the NLRIS doesn't change significantly. A significant change in NLRIS is observed only when either one among NLIQ or ICR is significantly larger than the other. This can be seen in the case of Trinity College Dublin where the difference in value between NLIQ and ICR is almost 0.4 and in the case of Uppsala University where the difference between NLIQ and ICR is close to 0.5. However,

Table 3: Correlation matrix that shows the correlation between NLIQ and ICR.

	NLIQ	ICR
NLIQ	1.0000	0.3205
ICR	0.3205	1.0000

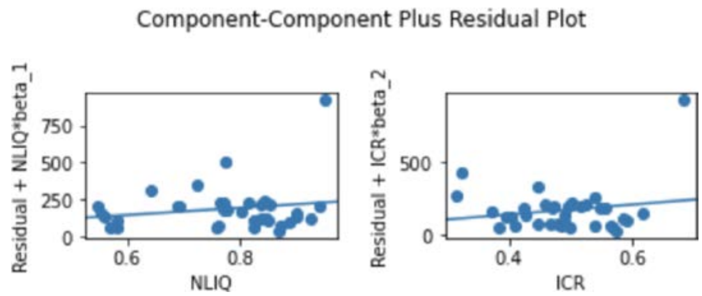

Figure 4: Regression graph of NLIQ and ICR VS QS Ranking(Q).

these anomalies are usually rare as any university which has a high global research influence is bound to attract many international collaborators and vice versa.

Results Analysis and Validation

Table 2 shows the NLRIS for 10 institutions in our data set. Multiple linear regression is used to check the significance of our research metrics on QS ranking. To start with, we checked for multicollinearity between the independent variables NLIQ and ICR. Table 3 shows the correlation between the non-local influence quotient and the international collaboration ratio. We see that the two metrics do not have a strong correlation and hence multiple linear regression can be performed.

Mathematically, the regression model is as:

$$Y = \beta_0 + \beta_1 \times X_1 + \beta_2 \times X_2$$

where Y = Internationality or NLRIS score, x_1 = NLIQ and x_2 = ICR. Figure 4, shows the results of the QS ranking regressed against ICR and NLIQ values. Furthermore, Table 4, shows the regression results along with the multiple for 0.01% and 0.05% confidence intervals. Table 4 shows the regression results for QS ranking regressed against the values of NLIQ and ICR. The model has poor adjusted R^2 value of 0.014. From Tables 5 and 6, we see that the confidence levels show a change of sign from negative to positive, clearly implying that variation in QS ranking can't be explained significantly by ICR or NLIQ at the specified confidence interval.

We have also plotted component residual plots, for the regression model to further strengthen our hypothesis. Figure 4 shows the relationship of each of the predictors (i.e NLIQ and ICR) to the dependent variable, QS ranking. From Figure 4, we can see that a reasonably strong relationship

Table 4: OLS Regression Results for QS ranking VS NLIQ and ICR.

Dep. Variable	QSranking	R-squared	0.069
Model	OLS	Adj. R-squared	0.014
Method	Least Squares	F-statistic	1.264
Date	Sat, 27 Mar 2021	Prob (F-statistic)	0.295
Time	22:45:04	Log-Likelihood	-238.31
No. Observations	37	AIC	482.6
Df Residuals	34	BIC	487.5
Df Model	2		
Covariance Type:	Non-robust		

	coef	std err	t		[0.025	0.975]
const	-163.982	225.520	-0.727	0.472	-622.296	294.330
NLIQ	243.538	231.714	1.051	0.301	-227.361	714.438
ICR	342.715	337.097	1.017	0.316	-342.348	1027.779

Omnibus	45.700	Durbin-Watson	1.119
Prob(Omnibus)	0.000	Jarque-Bera (JB)	202.396
Skew	2.815	Prob(JB)	1.12e-44
Kurtosis	12.979	Cond. No.	19.4

Table 5: Confidence interval 0.05% or QS Ranking VS NLIQ and ICR.

Variable Name	Confidence Interval	
const	-622.2958	294.3299
NLIQ	-227.3614	714.4380
ICR	-342.3476	1027.7786

Table 6: Confidence interval 0.01% or QS Ranking VS NLIQ and ICR.

Variable Name	Confidence Interval	
const	-779.2918	451.3260
NLIQ	-388.6692	875.7457
ICR	-577.0175	1262.4485

does not exist between our ranking parameters and the ranks awarded by QS ranking.

Next, we performed multiple linear regression to determine the amount of variation in NLRIS explained by NLIQ and ICR. Table 7 shows that both NLIQ and ICR are significant, and the amount of variation in NLRIS explained by ICR is more than NLIQ. We suspect that NLIQ may not follow the power-law relationship with citation count. We plotted citations vs NLIQ on a logarithmic scale (log-log plot). When the log-log plot shows a straight line, variables under study follow the power law. Figure 5 shows that citation count and NLIQ does not follow a power law, hence with the increase in citation count NLIQ does not necessarily increase, in an exponential fashion. This makes NLIQ very difficult to manipulate.

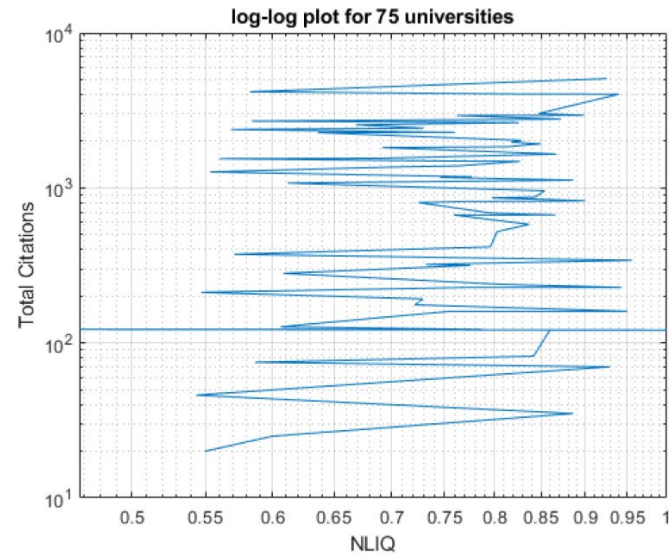

Figure 5: log-log plot for 75 Universities.

Table 7: Regression Results for NLRIS Vs NLIQ and ICR.

Variables	oefficient	std err	*t*	*P* > |t|	[0.025	0.975]
const	-0.0596	0.016	-3.784	0.000	-0.091	-0.028
NLIQ	0.4232	0.021	20.187	0.000	0.381	0.465
ICR	0.6960	0.020	35.122	0.000	0.656	0.735

DISCUSSION AND CONCLUSION

In this manuscript, our main motivation was to investigate, how the universities which are placed ahead by the popular ranking schemes perform when we evaluate them only based on the quality of research output, and specifically, global influence produced by their research output. We observe and analyze the following cases shown in Table 1.

Considering Brown University, its NLRIS is 0.53, relatively lower as compared to its QS rank which is 60. This university has 113 publications, out of which only 32 articles are written with international collaboration. Also, these 113 articles have received 2381 citations with 1027 citations from the university's own country. The ICR is 0.49 and NLIQ is 0.57 hence a lower NLRIS value. But the university is ranked high, 60 in the QS rankings. Considering 'University of Helsinki' which is ranked lower than the Brown University in QS rankings is given high NLRIS by our model. The University of Helsinki' has a 115 publication count, with 52 articles written with international collaboration. Also, out of 2778 citations, only 356 citations are from its country leading to a high NLIQ score of 0.87. QS rankings consider 4 components, academic reputation, employer reputation, citations per paper, and *h*-index which accounts for public perception along with research output. It is noticeable that QS

Table 8: Universities not featured in QS ranking: NLRIS and metadata.

Institution	No of articles	Articles with global collaboration	Total citations	Parent country citations	NLRIS
Atilim Uniersity	31	6	416	85	0.7134
University of Lausanne	38	11	688	140	0.6795
University of Munster	77	20	2025	348	0.6329
University of St Andrews	36	18	819	136	0.5908
University of California Santa Cruz	95	26	2274	990	0.5858

ranking or other ranking schemes that consider the learning environment, volume of research, income, and reputation, statistics of students, staff, etc. does not cater to the evaluation of the non-local influence of the institutions.

We propose the evaluation metric which is fair to the institutions which are smaller in size and do not qualify to be considered for the reputed ranking schemes but are performing well in terms of research quality and attracting citations and collaborators around the world. In our data set, we analyze a few universities like Atilim University, University of Lausanne, University of Munster, University of St Andrews, University of California Santa Cruz in terms of their research output in the domain of Computer Science. These universities do not appear in the Computer Science QS ranking. Table 8 shows the number of articles published by these universities along with the amount of international collaboration and non-local citations.

Our metrics are size-independent and hence are not biased towards the universities big in size, diverse, having larger intakes, funding's, good student-staff ratio, academic and employer reputation, public perception, etc. It is noticeable, that even with less volume of publications, these universities are receiving citations and attracting collaborators around the world and not just from the local group within their own country. Hence the standard notion or public perception that the universities big in size also do well on the research front may not always hold if we focus on very specific research-based metrics.

In their work, authors proposed ALIS (Author Lineage Independence Score), an influence measurement method for scholars using the Geneology tree, and showed that the ALIS metric does not follow preferential attachment.[12] Authors in their work proposed a mathematical model called BA model of network growth and introduced 'preferential attachment'.[13] They showed that new nodes tend to get attached with old nodes with high in-degree. The resulting networks follow a power law. We must note that the *h*-index in the citation network does follow a power law, and therefore for established authors, the citation pattern, as reported by *h*-index, gets a boost to new citations, not necessarily reflective of the quality of the cited article. So the rich become richer. We have verified that NLIQ does not follow the BA model/power law. Hence with the increase in citation count, NLIQ will not increase unless there are a good proportion of citations coming from outside of the home country. This establishes NLIQ as a robust quality indicator, particularly for institutions in disadvantaged locations or for those institutions not able to boast of 'favourable' perception.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

REFERENCES

1. Aithal PS, Kumar PM. ABC Model of Research Productivity and Higher Educational Institutional Ranking. International Journal of Education and Management Engineering. 2016;6(6):74-84. doi: 10.5815/ijeme.2016.06.08.
2. Madhan M, Gunasekaran S, Arunachalam S. Evaluation of research in India - are we doing it right?. Indian J Med Ethics. 2018;3(3):221-9. doi: 10.20529/IJME.2018.024, PMID 29650499.
3. Docampo D, Bessoule JJ. A new approach to the analysis and evaluation of the research output of countries and institutions. Scientometrics. 2019;119(2):1207-25. doi: 10.1007/s11192-019-03089-w.
4. Massucci FA, Docampo D. Measuring the academic reputation through citation networks via PageRank. Journal of Informetrics. 2019;13(1):185-201. doi: 10.1016/j.joi.2018.12.001.
5. Nykl M, Ježek K, Fiala D, Dostal M. PageRank variants in the evaluation of citation networks. Journal of Informetrics. 2014;8(3):683-92. doi: 10.1016/j.joi.2014.06.005.
6. Kazi P, Patwardhan M, Joglekar P. Towards a new perspective on context based citation index of research articles. Scientometrics. 2016;107(1):103-21. doi: 10.1007/s11192-016-1844-2.
7. West JD, Jensen MC, Dandrea RJ, Gordon GJ, Bergstrom CT. Author-Level Eigenfactor Metrics: evaluating the influence of authors, institutions and countries within the SSRN community. Harvard Business School NOM unit working paper. SSRN Journal. 2012:(12-068). doi: 10.2139/ssrn.1636719.
8. Abramo G. Revisiting the scientometric conceptualization of impact and its measurement Journal of Informetrics. 2018;12(3):590-7. doi: 10.1016/j.joi.2018.05.001.
9. Ginde G, Saha S, Mathur A, Venkatagiri S, Vadakkepat S, Narasimhamurthy A, Daya Sagar BS. 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. Bodenstein C, Schryen G, Neumann D. Energy-aware workload management models for operation cost reduction in data centers. European Journal of Operational Research. 2012;222(1):157-67. doi: 10.1016/j.ejor.2012.04.005.
11. Kennedy J, Eberhart R. Particle swarm optimization. In: Proceedings of the ICNN'95-international conference on neural networks. Vol. 4. IEEE Publications; 1995.
12. Dey S, Mathur A, Saha S, Bhattacharya S, Dayasagar B. ALIS is wonderful: An experimental study in Lineage Independent Influence Evaluation of Scholars; unpublished.
13. Barabási AL. Network science. Available from: http://networksciencebook.com/chapter/5barabasi-model [cited 23/3/2022].