Journal of Scientometric Res. 2022; 11(3):309-317
http://www.jscires.org

Research Article

How Effective is Research Funding? Exploring Research Performance Indicators

Saira Hanif Soroya[1,2,3,*], Muhammad Umar[3], Naif R. Aljohani[4], Anna Visvizi[5,6], Raheel Nawaz[7]

[1]University of Kentucky, Lexington, USA.
[2]University of the Punjab, Lahore, PAKISTAN.
[3]Department of Computer Science, Information Technology University, Lahore, PAKISTAN.
[4]Faculty of Computing and Information Technology, King Abdulaziz University, Jeddah, SAUDI ARABIA.
[5]SGH Warsaw School of Economics, Warsaw, POLAND.
[6]Effat College of Business, Effat University, Jeddah, SAUDI ARABIA.
[7]Staffordshire University, Staffordshire, UNITED KINGDOM.

ABSTRACT
This paper deploys bibliometric indices and semantic techniques for understanding to what extent research grants are likely to impact publications, research direction, and co-authorship rate of principal investigators. It includes semantic analysis in the research funding evaluation process to effectively study short-term and long-term funding impact on publication outputs. Our dataset consists of researchers who received research grants from the National ICT Research and Development funding program of Pakistan. Whereas Pakistani researchers' publications dataset was extracted from Scopus. We show several interesting case studies to conclude that bibliometric-based quantitative assessment combined with semantics can build better sustainable pathways to deploy evaluation frameworks for research funding effectively. The funding data of closed projects from 2007 to 2013 was obtained from ICT R&D public records. The publications dataset was extracted from Scopus data and the details of the statistics were, publications=61,421; researchers=42,376, organizations=213; funded projects=17, funded researchers=23 and funded organizations=10. A significant positive impact (more research output after allocation of funds) has been found for almost all studied organizations. Similarly, a positive funding impact on research output and average co-authorship for the studied cases (investigators under consideration) was found. However, no funding impact was found on the research focus of investigators, i.e., research focus remained almost unchanged after grant allocation. Also, the study suggested the best possible match candidates for collaboration or potential reviewers against the selected project by semantically analyzing the executive summary. Most funded researchers and research organizations have found a positive funding impact on research output (i.e., number of publications). Using semantics along with bibliometric indicators (relating to funding and impacts) can be constructive in making funding programs more effective and for better impacts evaluation; it is recommended for funding agencies to use it in formal framework formation and proposal evaluation process.
Keywords: Research Evaluation, Research Grants, Sustainable Impact, ICT R&D Fund, Bibliometrics.

Correspondence
Saira Hanif Soroya
[1]University of Kentucky, Lexington, USA.
[2]University of the Punjab, Lahore, PAKISTAN.
[3]Department of Computer Science, Information Technology University, Lahore, PAKISTAN.
Email id: sairasroya@gmail.com
ORCID ID: 0000-0002-8153-1529

Received: 19-06-2022;
Revised: 27-09-2022;
Accepted: 07-11-2022.
DOI: 10.5530/jscires.11.3.34

INTRODUCTION

Research and Development (R&D) are the backbones of the globally competitive knowledge-driven economies. The investment in R&D is significant for the global economies to develop new products and services that drive growth, create jobs, and improve the sustainable smart city growth.[1-3] However, scientific research is becoming very expensive due to the interdependence of each branch with other disciplines.[4,5] Due to the ever-increasing cost of competitive research, many countries are primarily finding a solution through the joint funding of laboratories and projects for sustainable outcomes,[6-9] e.g., CERN[1] is the largest particle physics laboratory in the world which is operated and funded by countries in European Union (EU).

While the developed world can afford the luxury of many research institutes and centres, the situation in the developing world is alarming. According to Dehmer *et al*.[10] concerning the global R&D expenditures, the developed countries are participating a lot more in research compared to developing countries. The United States is currently the global leader in R&D spending, having almost one-third portion of global

1 https://home.cern

R&D spending, followed by China and Japan; however, if the unions of countries are also considered as combined units of spending, the EU then becomes second, followed by China as a third major contributor to global research.[11]

Recently, knowledge has become crucial for global economies since many economies have become knowledge bases.[12,13] Many countries are striving for their economies to be knowledge-based and so spending a vital part of their Gross Domestic Product (GDP) on R&D. By the 1960s US accounted for almost 70% of the world's R&D funding. However, in 2019 the rest of the world accounted for 70% of R&D funding, and the US got a 30% share. It is not because the US has reduced investment in R&D but because the rest of the world is seriously investing in the field of R&D.[14] Israel, Finland, and Qatar are tiny (i.e. area and population-wise) countries but spend a higher percentage of their GDP on research in contrast to some other big countries and economies (e.g. Brazil and India). Let us look into top global economies' current growth rates and investment in R&D. China's total funding for R&D is expected to surpass that of the US by about 2022, which indicates higher growth in R&D spending by China as compared to the US.[15] However, a question remains about how the success and impact of R&D activities are measured? So far, publishing technical papers with maximum citations, patents, and new product introductions is a key measure of success.

Among the developing nations, the research landscape of Pakistan has been very impressive over the last decade and shows continuous growth.[16] Research trends are improving, and initiatives have been taken to promote innovation through research projects and ideas. Based on the research output statistics, Pakistan is expected to be the second-fastest-growing country in research output, just after Malaysia.[17] Due to scarce resources (e.g., funds) in Pakistan and the ever-increasing cost of carrying out cutting-edge research, it is hence necessary to make the best use of funds and scientific manpower to achieve high scholarly impact. It thus becomes important to develop a formal understanding of the key factors that influence the relationship between funding, research output, and impact and to develop methods to meaningfully measure the impact of research funding.[18]

In this paper, we studied the relationship between research funding and research output and impact with the objective of developing a formal analytical framework. We employed metrics for evaluating various dimensions of research output and impact for the accurate measurement of national research productivity. An important aspect of this study is the examination of the longer-term impact of funding beyond the grant period - which provides information about funding impact that goes beyond the time embodied in the final reports commonly submitted at the end of a grant. We examined Pakistani researchers who received funds from the National ICT R&D fund program – a flagship national research-funding program that aims "to transform Pakistan's economy into a knowledge-based economy by promoting efficient, sustainable and effective ICT initiatives through synergic development of industrial and academic resources."[2]. The study objectives are as follows:

Analysis of performance indicators relating funding to research productivity and impact.

Studying research productivity per unit of funding at the organization level and the individual researcher level.

Long-term sustainable impact of research funding on research activity beyond the duration of the grant by deploying semantic analysis on the scientific literature published by the scholars.

The novelty of this paper lies in the fact that a lot of work has been done by deploying various bibliometric indicators to better describe the relationship between researcher, research, funding, and the impact of research funds; but this work is among the few studies that use semantics to evaluate the scientific impact of research. The rest of the paper has been structured as follows: Section 2 includes previous work related to methodologies of impact assessment of funding, data and text mining, the importance of semantics in revealing trends, and some discussion on the challenges in assessing impacts of funding. Section 3 describes the datasets, data collection procedures, and analytical techniques to measure similarity and closeness among documents. Section 4 includes all the major analyses and measures carried out during this study. Section 5 presents a case study to illustrate the use of our deployed measures to study the impact of the ICT R&D Fund project. Finally, Section 6 presents conclusions and makes some recommendations for future work.

Literature Review

This section reviews the related literature from the following two perspectives: In the first half, we presented a brief review of the performance indicators relating funding to research productivity; in the second part, we presented a brief review of related semantic and text mining techniques used to study the long-term impact of research funding on the scientific literature published by the scholars.

Review of research performance indicators

The research and Development (R&D) funding programs play a vital role in public research policy, and these schemes lead toward improving quality research output. The sole purpose located beneath these programs is to make sure that competitive grants are helping to enhance research

2 http://www.ictrdf.org.pk/

performance and to get the most of it. Studying the impacts of such funding schemes has become more common in such a way that it makes funding authorities more curious and keener to make sure that these grants have intended positive impacts on research performance and scientific quality.[19]

Several studies examine the relationship between funding and research output at the level of the university or department; some examine international scientific knowledge flows and the scholarly impact of countries and institutions.[20-22] Other studies we came across examined the funding-output link at the level of the individual researcher and found small positive effects.[23,24] A few studies provide impact assessment for progressing research, and some work on assessing the economic impact of R&D.[25,26] Results of these studies show higher increases in the number of publications for grant recipients than for rejected applicants, while increases in mean normalized citation rates were not significantly higher for the successful applicants.[27,28] However, it should also be noted that these productivity increases also include greater productivity of highly cited papers.[19]

A few studies have focused on inspecting the connection between researchers' past performance and the peer review of grant proposals and found that the application candidates have a tendency to have better track records in the form of higher citation scores than non-applicants and rejected candidates.[29-31] Similarly, Chilean research funding found a noteworthy impact on several publications (i.e., higher research output for successful candidates in contrast to rejected candidates), but no significant impact was found on citations.[32] Furthermore, Jacob and Lefgren[33] carried out a study on NIH postdoc grants impacts. They found a positive impact on research output for granted applicants, i.e., the number of publications increased with the reception of grant, also increase in the probability of crossing a citation threshold has been noted, but a total number of citations got no impact on successful and unsuccessful applicants. Melin and Danell[34] showed that young investigators in Sweden whose applications were selected for a 6-year grant didn't impact the number of publications as expected (no improvement for awarded applicants in research output as compared to rejected ones); however positive impact has found in terms of international co-authorship, which helped the research groups for securing furthers funding.

Recently, Langfeldt *et al*.[35] studied the impact of Norway's FRIPRO and Denmark's DCIR funding agencies and found a significant impact on the funding schemes in both countries. The higher increase in research output and highly cited papers had seen successful applicants in contrast to rejected candidates. However, did not find any notable impact on average citations for successful and rejected applicants, concluding no impact on the importance of research.

Review on text mining and semantics

Data mining and text mining approaches can be utilized to find patterns from structured data.[36,37] These techniques can be applied to unstructured scholarly big data corpus to convert it into structured data to reveal valuable insights. Text mining can also be used to discover and extract patterns and trends in scholarly data by deploying Natural Language Processing (NLP).[38] The NLP is a technique that is widely used to extract information from textual data. NLP is particularly effective in the extraction of predefined patterns or existing information that can help in finding and exploring trends from any database.[39]

An investigation on document similarity found that commonly used similarity techniques such as the cosine and Jaccard treat the words as independent entities from one another, which is, however, unrealistic as words in documents combine to deliver proper context. Words in any document are interrelated to form meaningful structures and develop ideas. An alternative is suggested to use concepts instead of words to extract the topics of documents by resolving redundancies (i.e., synonymy) and ambiguities (i.e., polysemy) in words.[40]

Another study also describes the same weaknesses (i.e., treating words as independent entities) of the BOW (Bag of words) approach used in common similarity measures and suggests an alternative to overcome the situation and get improved accuracy. BOW typically represents the text in the vector space model and does not consider redundancies (i.e., synonymy) and ambiguities (i.e., polysemy) problems and ignores semantic relatedness among words. To overcome the shortages of the BOW approach, the suggested alternative is to embed WebNet-based semantic relatedness measure for pairs of words into a semantic kernel. This measure incorporates the TF-IDF weighting scheme; thus, semantic and statistical information is combined from the text to provide improved classification accuracy.[41,42] Semantic tools are gaining popularity in assessing the efficiency of funding programs.[43]

In this study, we used the vector-space model to represent text obtained from the scientific document. Thus, the text is represented by the vectors of terms extracted from the documents; associated weights are then assigned to define the importance of these terms in respective documents and the collection of documents under consideration. The weights (see Equation 1) of terms are normally calculated using the TF-IDF method, in which the weight of a term is determined by two factors: how often the term j occurs in document *i*, i.e. the term frequency $tf_{i,j}$ and how often it occurs in the whole collection of documents, i.e. the document frequency df_j.[44]

$$W_{i,j} = tf_{i,j} \times \log\left(\frac{N}{df_i}\right) \tag{1}$$

After term weights are determined, the ranking function is used to measure the similarity between the query and document vectors. Cosine measure is a commonly used similarity measure that is used to calculate the angle between the query vector and documents vector when they are represented in a V-dimensional Euclidean space, where V is the vocabulary size,[45] see Equation 2.

$$\mathrm{sim}(Q,D_i) = \frac{\sum_{j=1}^{v} W_{Q,j} \times W_{i,j}}{\sqrt{\sum_{j=1}^{v} W_{Q,j}^2 \times \sum_{j=1}^{v} W_{i,j}^2}} \quad (2)$$

We presented a brief review of the related literature on the performance indicators relating funding to research productivity. In addition, we also discussed techniques to semantically analyze the data using text mining and clustering techniques. While performance indicators employed in this study include publications/unit of funding, research output growth for Principal Investigators (PIs), Principal Investigators Organizations (PIOs), growth in researcher co-authorship rate, the semantic analysis is carried out to study the long-term impact of research funding on the scientific literature published by the scholars for assessing research focus or scholars and long-term funding impact. Combining both statistical bibliometric and semantic analysis for performance/ impact measures leads to a better understanding relationship between research output and research funds.

DATA AND METHODS

Text mining techniques, including information retrieval and data mining, have been used to carry out the analysis. In addition, a web scraping technique has been used to automate the data retrieval process of ICT R&D-funded projects data, and wrappers/parsers have been developed to extract formatted data from Scopus based on all Pakistani research publications from CSV dataset. Furthermore, text processing, vector space modelling, and similarity/distance measures have been employed to find out similarities among documents. Bibliometric analysis has been carried out to assess the impact between funding, proposals, researchers, organizations, and publications.

Dataset

To carry out the analysis, two types of datasets were needed, i.e. funding agencies funded projects data and researchers' publications output data. The first dataset includes the information on funded projects by the ICT R&D funding agency of Pakistan, available on their website[3]. The second dataset belongs to Pakistani researchers' publications data downloaded from Scopus. The datasets include information about Principal Investigators (PIs), Principal Investigators Organizations (PIOs), abstract or summary of proposal and output, funding amount, duration and year of grant, and publications.

The funding data, obtained from ICT R&D public records, made limited to only closed funded projects from 2007 to 2013. These are the projects that were funded and are completed successfully and have some output available for long-term impact evaluation. We did not include information on ongoing funded projects in our dataset because that might mislead results, as it doesn't include required attributes for analysis.

All Pakistani researchers' publications dataset was extracted from Scopus data. It was then pre-processed to exclude unnecessary attributes and filtered to include only publications during 2005-2013 for the sake of simplicity and was kept in CSV format for further processing and analysis. This dataset includes information from more than sixty thousand publications (output of around forty thousand unique researchers that belong to more than two hundred distinct organizations). The followings are the statistics of data used in this study: publications= 61,421; researchers=42,376, organizations=213; funded projects= 17, funded researchers = 23 and funded organizations =10.

Methods

To evaluate the performance of the overall research funding program, funding agencies ought different impact evaluations of their funding schemes to see their effects on research performance and scientific quality. Our approach to evaluating funding impacts has been divided into four stages. These stages are 1) grant allocation, 2) research output, 3) research collaboration, and 4) long-term research impacts.[46]

During *the grant allocation* stage, the funding recipients are characterized, and analyses of funding allocation patterns are carried out as a method for assessing the impacts of research activities. This quantitative data can then further be utilized in other comparisons and analyses, e.g., comparisons between funding recipients and rejected candidates, etc. In the second stage of *research output* evaluation, an analysis of scientific output related to funding is carried out, i.e., calculate the number of publications produced, as well as identify factors that influence the output volume, e.g., funding duration, research team, etc. The main focus of the analysis was to find out opportunities and outcomes brought about by research funding (see Equation 3).

$$\text{productivity}_{pub} = \text{avg(publications)}_{ag} - \text{avg(publications)}_{bg} \quad (3)$$

where pub=publications,ag=after grant,and bg=before grant

3 http://www.ictrdf.org.pk/

The third stage includes the impact of research funding on research users; which is not linear and hard to identify and quantify; so, *research collaboration* is assumed as a proxy measure for impacts evaluation because it has a positive impact on the performance of the innovation system. The analysis includes research collaboration impacts evaluation of funding by examining increased or decreased average co-authorship rate after funding (see Equation 4).

$$\text{impact}_{\text{co\_auth}} = \text{avg}(\text{co\_auth rate})_{\text{ag}} - \text{avg}(\text{co\_auth rate})_{\text{bg}} \quad (4)$$

where co_auth=co authorship,ag=after grant, and bg=before grant

The final stage deals with *long-term impact* evaluation; analysis focusing directly on the impacts on research users is carried out. The assessment of such types of impacts is very challenging because it includes employing both qualitative and quantitative approaches to measure aggregated impacts. These impacts include evaluation of research direction focus, productivity and other indicators of individual researchers with respect to funding (see Equation 5).

$$\text{impact}_{\text{rf}} = \text{avg}(\text{own similarity})_{\text{ag}} - \text{avg}(\text{own similarity})_{\text{bg}} \quad (5)$$

where rf=research focus,ag=after grant ,and bg=before grant

Also, during this phase, most appropriate candidates' suggestion is made against any research proposal based upon the candidate's publications similarity with respect to the research proposal summary. This analysis helps to evaluate the researcher's expertise in the area he/she has applied for funding compared with all the researchers in the dataset. The outcomes of this analysis can also be used to find appropriate reviewers working in related areas. The candidates with higher/maximum publications similarity with the research proposal are shown as Equation 6.

$$\text{candidate suggestion} = \max\left(\text{avg}_{i=1}^{m}\left(\text{sim}_{j=1}^{n}(P, Q_{ij})\right)\right) \quad (6)$$

The applied document similarity algorithm in Equation 6 was broken into three major parts; pre-processing of documents, document's text to vector space model, and documents similarity measure. For this process, documents are passed through a pre-processing phase which involves first tokenizing the documents, then these tokens are transformed into the same casing process, e.g. lower casing, and finally, stop-word removal filters and stemming processes are applied. After pre-processing of documents, it is then mapped to a vector, and weights are assigned to each term on the basis of TF-IDF, which assigns weights based on term importance and its occurrence in the document collection. After assignments of weights, the final step is to calculate the similarity between documents (project or publication summary); for this purpose, the cosine similarity measure has been taken into account, which is quite a commonly used measure. The cosine similarity measure is used to calculate the similarity between documents based on the weights of the words represented by the two document vectors, as shown in Equation 7.

$$W_{i,j} = tf_{i,j} \times \log(N/df_i) \quad (7)$$

The result of cosine similarity is always between zero and one. While zero indicates no similarity between documents, one indicates that documents are the same. The high similarity output indicates that documents are more similar and closely related to each other, whereas low similarity results indicate that the documents are less similar and more different. The formula to compute cosine similarity is presented in Equation 8.

$$\text{Similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\|\|B\|} = \frac{\sum_{i=1}^{n} A_i \times B_i}{\sqrt{\sum_{i=1}^{n}(A_i)^2 \times \sum_{i=1}^{n}(B_i)^2}} \quad (8)$$

RESULTS AND DISCUSSION

In this section, we deploy bibliometric performance indicators and semantics to understand PIs' research focus, diversity, productivity, and co-authorship impact on w.r.t funding. The purpose of these analyses is to measure the research productivity of PI after the grant is received, co-authorship rate (research collaboration), and research focus related to the theme of the received grant. The data is divided into two periods, one before (excluding application year) and one after (from funding year onwards) the funding decision. The differences between the two periods are analyzed based on the following indicators: a) Researchers' average number of publications per year; b) Researchers' average co-authorship rate per year; c) Researchers' research focus w.r.t. funded theme per year and d) Research organizations' funding and productivity statistics.

First of all, the funding agency's data is selected for carrying out analysis. The purpose of these analyses is to see the trends in funding data, to which the funds are allotted, agency's yearly grants, yearly funded projects, number of researchers per project, average funding per project, researcher-wise grants, and organization-wise grants, and years in which maximum and minimum grants are allocated, etc. The funding dataset includes researcher names, proposal summary, grant information, year of the grant, and status of the project i.e. closed or in progress. This study is carried out for only funded projects which are in closed status because one intention of this study is to investigate the long-term impact of funding.

Secondly, all Pakistani research publications data is utilized for analysis to see the behavior of data i.e. researcher's total and yearly publications, organizations' total and yearly publications, publications' average co-authorship rate, etc. These analyses, when combined with funding agencies' data, become very helpful in understanding the flow of funding as well as researchers' and organizations' research trends concerning grants. Funding impacts on research organizations have also been assessed during this study but are made limited to only productivity analysis (increase or decrease in research output) for the sake of simplicity and limitations of scope (future works can cover other indicators including quality of research etc.). Figure 1 lists all the funded organizations along with the information on total assigned projects by the funding agency, total allotted funds by the agency, and the productivity (in terms of the average increase in publications w.r.t grant). All organizations depict positive impacts of funding on their productivity. For this analysis, funded organization data is obtained along with their names mapping (funding agency's mentioned organization names and Scopus dataset used organization names) and then publication statistics are extracted from the Scopus dataset for selected organizations. Note that the funds are shown in PRK (104 PRK ~ 1 USD), and the names of organizations were kept anonymous.

Impacts of funding on researcher's collaboration are also evaluated based on average growth (positive or negative) in co-authorship rate comparing two periods of publications i.e. prior and posterior to grant. Other semantic analyses include assessing the impacts of funding on research focus and the diversity of researchers. For this purpose, each researcher's publications are semantically analyzed (similarity measures are carried out among the researcher's publications) to see how much specific an individual researcher is in his/her research area. The higher similarity shows more area-specific research, and less diversity and vice versa. Figure 2 shows the results of such analysis for two researchers where they are mapped along with their co-authorship rate impact (increase or decrease in average co-authorship rate after allocation of funds) and research focus impacts (positive or negative impact comparing

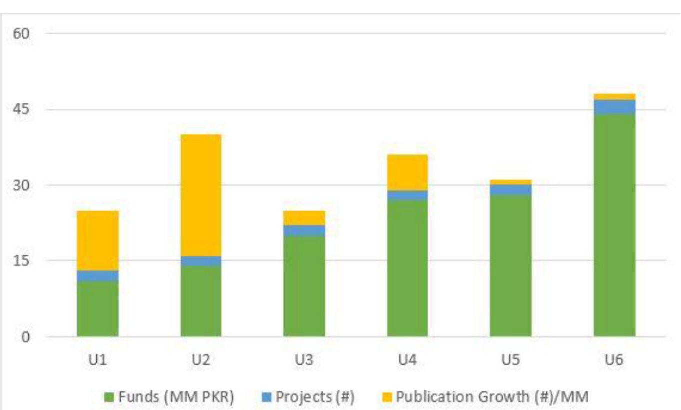

Figure 1: Funded organizations' productivity w.r.t funding in PKR.

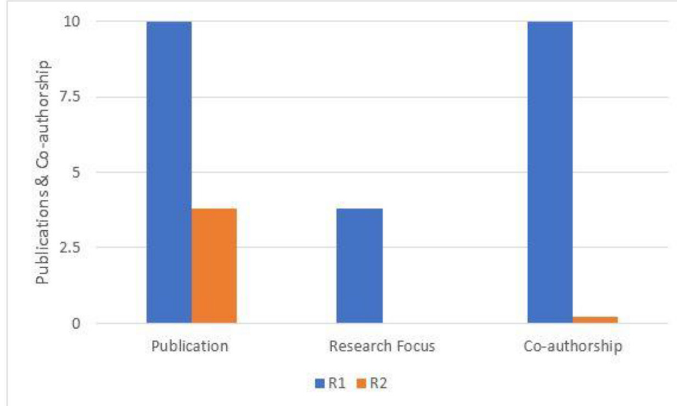

Figure 2: Funding impact on funded researchers' publications and co-authorship.

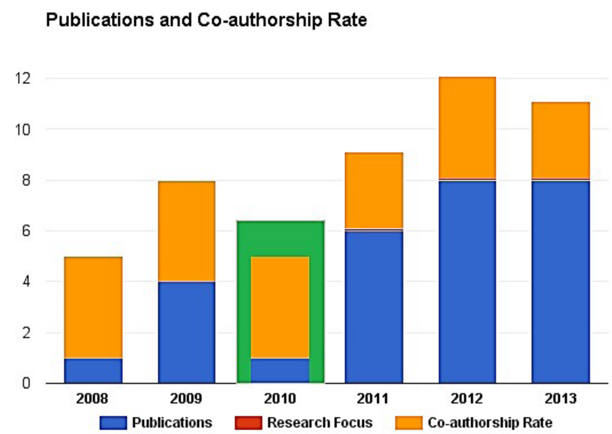

Figure 3: Funded researcher's (Case 3) yearly productivity funding impacts.

before and after grant periods). Higher positive funding impacts on research focus and co-authorship can be seen for PI (Researcher 1) in comparison with P2 (Researcher 2); having no funding impact on research focus means the P2's (Researcher 2) research focus did not shift towards the theme funded was granted and very low positive impact on co-authorship.

Another analysis on funded researchers at the individual level is carried out to see the researcher's yearly progress and funding impacts. Figure 3 shows that a funded researcher (Researcher 3) whose yearly productivity (increase or decrease in publications count) and impacts (average increase or decrease in co-authorship rate and research focus) before and after the grant are mapped and the funding year which is 2010, in this case, is marked as green background. Positive funding impacts on publications growth and research focus are found in this case, however; research focus has observed a slightly negative funding impact.

Interesting results were observed during the analysis phase. In an analysis of organizations' allocated funds and research

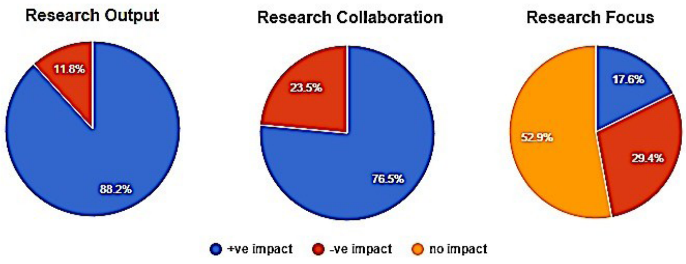

Figure 4: Researchers' funding impacts evaluations.

output, a positive impact (more research output after allocation of funds) has been found for almost all organizations. However, some organizations have been found more productive than others in terms of higher publication rates per unit of the fund. While analyzing researchers' productivity and impact measure w.r.t funding, the positive impact has been seen for most of the researchers, i.e. higher research output rate (positive difference) found comparing two periods i.e. before and after grant. However 'No' or negative impact (equal or less research output rate after allocation of funds) of funding has also been observed in some cases. Hence, researchers' performance is on average better after the grant. In examining the impact of funding on the co-authorship rate of researchers, again the positive impact of funding has been seen for most of the cases and negative impact for very few cases; depicting the overall positive funding impact on researcher's co-authorship rate. However, mixed results have been found while measuring the researcher's research focus and diversity of research, before and after the grant, i.e. in most of the cases no funding impact has been found, and positive and negative impact have found for the rest of the cases; overall resulting no significant funding impact on researcher's research focus. Evaluations of the impacts and productivity analysis can be seen in Figure 4.

Suggested candidates (potential reviewer or collaborator) analysis has also been performed against each funded project to list down the best possible match candidates (based upon their higher research similarity with the project and higher publications average) to whom the funding agency can coordinate with, during the proposal evaluation process, funding allocation process and/or ongoing project assistance.

Case Study: Measuring the impact of National ICT R&D-funded project

We present a case study to illustrate the use of our deployed measures to analyze the impact of an ICT R&D-funded project. For this purpose, we selected a project out of available seventeen closed projects funded by the ICT R&D funding agency from 2007-to 2013. We then analyzed principal investigators (PIs) and principal investigators' organization (PIO) productivity and funding impacts of the selected project; these analyses were carried out on all Pakistani research output data extracted from Scopus from 2005-to 2013. In addition, we suggest the most appropriate list of candidates who could be potential reviews or collaborators having higher research relevance to the selected project. Finally, we discuss the results of the semantic and bibliometric analysis of the selected project.

We selected a successfully closed funded project to analyze funding impacts on the researcher, organization, and research work itself. The information required to carry out the necessary analysis included project title, project summary, principal investigator's name, principal investigator organization, funding year, and allocated budget (Table 1). This project was funded in 2009 and our dataset contains research information from 2005 to 2013, so long-term impacts can be easily assessed.

We analyzed the selected organization's productivity to see whether either selected organization has observed any positive impact of funding in terms of research output growth. The organization's average increase in research output statistics is mapped along with their assigned funded projects and grants by the ICT R&D funding agency. Positive funding impact on an organization's publications growth i.e., around three publications per million PKR can be seen in Figure 1 (appendix).

We then analyzed funding impacts at the individual researcher level. Principal investigators are assessed based on average publications growth, increase or decrease in research focus and positive or negative change in average co-authorship rate before and after funding. Figure 2 (appendix) depicts

Table 1: Descriptive information of selected funded project.

Project Title	ABC
PIs	Researcher 1 (R1) & Researcher 2 (R2)
PIO	University 1 (U1)
Year	2009
Budget	PKR 13.03 million (~ 130k USD)

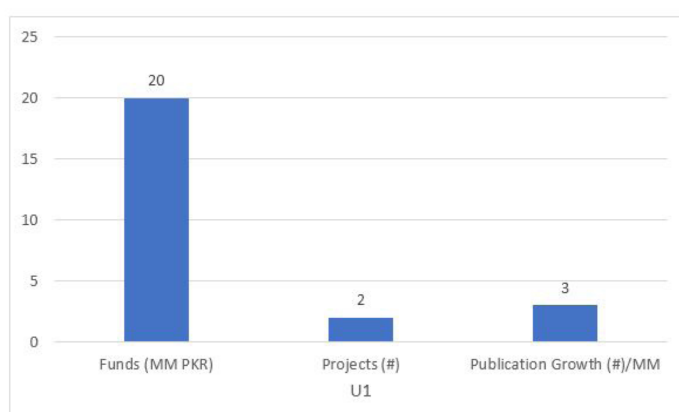

Figure A-1: Organization's funding and productivity statistics 2005-2013.

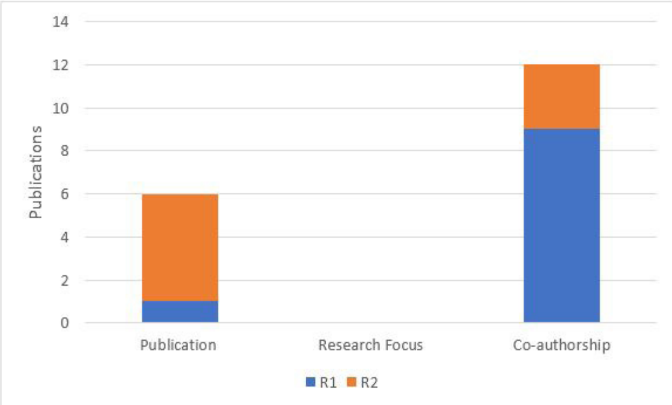

Figure A-2: Principal investigators' funding impacts statistics.

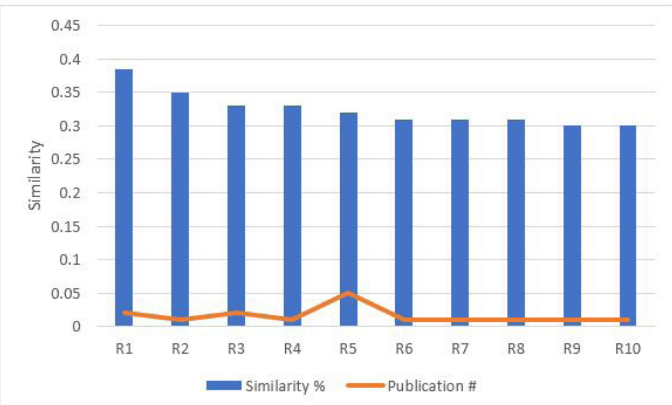

Figure A-3: Candidates suggestions based upon research similarity.

a positive funding impact on research output and average co-authorship for both the investigators under consideration; however, no funding impact was found on the research focus of investigators, i.e., research focus remained almost unchanged after grant allocation.

Finally, we suggest the best possible match candidates for collaboration or potential reviewers against the selected project by semantically analyzing the executive summary. The project's executive summary is matched with each researcher's publication abstract in the dataset. Finally, the researchers with similar fields are extracted and then highly similar researchers are suggested. In Figure 3 (appendix), suggested researchers are mapped along with their research output statistics. We find one of the PI (R 8) appears in the list, suggesting that the researcher has expertise in the field to execute this project.

Concluding remarks

This is a novel approach toward impact assessment that deals not only with the bibliometric analysis to evaluate the impacts of funding on research but also deploys semantic analysis to make this evaluation more appropriate and accurate. This study examines the measures, period before funding and after – the difference between before and after the grant is received. Positive funding impact on research output (i.e. number of publications) has been found for almost all funded researchers and research organizations. The approach of using semantics along with bibliometric indicators (relating to funding and impacts) can be very helpful in making funding programs more effective and for better impacts evaluation; it is recommended for funding agencies to use it in formal framework formation and/or proposal evaluation process.

Similarity-based analysis during this study is carried out using cosine similarity measure, which is very simple and efficient but treats the words independently; thus, it does not include relatedness between the words, which may lead to poor results. Better and more accurate results can be achieved by embedding semantic relatedness[47-49] between words and resolving redundancies and ambiguities, as examined in the following studies.[50-53] Future works can also include suggesting funding agencies for currently hot research areas the rest of the world is working on using deep learning methods.[54-57] These research trends evaluation, when combined with funding impact evaluation, might lead to better framework development for funding agencies which covers the processes of area selection, proposal evaluation, ongoing research assistance, and finally, the long-term sustainable impact of research assessment.

REFERENCES

1. Khalid S, Hassan SU, Sitthi A. Towards Smart Data Management of Scientific Literature: Addressing Polysemy and Aberrant Decoding in Author Names. In The International Research and Innovation Forum. Springer, Cham. 2021:435-45.
2. Hassan SU, Haddawy P, Zhu J. A bibliometric study of the world's research activity in sustainable development and its sub-areas using scientific literature. Scientometrics. 2014;99(2):549-79.
3. Saeed-Ul-Hassan PH, Kuinkel P, Sedhai S. A bibliometric study of research activity in sustainable development. In Proceedings of 13th Conference of the International Society for Scientometrics and Informetrics. 2011;2:996-8.
4. Kayani M, Hassan SU, Aljohani NR, Dancey D, Liu L, Nawaz R. Towards Interdisciplinary Research: A Bibliometric View of Information Communication Technology for Development in Different Disciplines. In 2019 Third World Conference on Smart Trends in Systems Security and Sustainability. IEEE. 2019;185-92.
5. Hassan SU, Ichise R. Discovering research domains using distance matrix and co-authorship network. In SDM. 2009;3:1252-7.
6. Hassan SU, Muazzam A. Analyzing institutional research publications, citations and collaboration activities in sustainable development and its sub-areas, with particular reference to Asian institutions. Journal of Sustainability Science and Management. 2013;8(2):180-90.
7. Tuarob S, Assavakamhaenghan N, Tanaphantaruk W, Suwanworaboon P, Hassan SU, Choetkiertikul M. Automatic team recommendation for collaborative software development. Empirical Software Engineering. 2021;26(4):1-53.
8. Hassan SU, Iqbal S, Aljohani NR, Alelyani S, Zuccala A. Introducing the 'alt-index for measuring the social visibility of scientific research. Scientometrics. 2020;123(3):1407-19.
9. Drongstrup D, Malik S, Aljohani NR, Alelyani S, Safder I, Hassan SU. Can social media usage of scientific literature predict journal indices of AJG, SNIP and JCR? An altmetric study of economics. Scientometrics. 2020;125(2):1541-58.
10. Dehmer SP, Pardey PG, Beddow JM, Chai Y. Reshuffling the global R&D deck. PloS one 2019;14(3):1980-2050. e0213801.
11. Bonaccorsi A, Cicero T, Haddawy P, Hassan SU. Explaining the transatlantic gap in research excellence. Scientometrics. 2017;110(1):217-41.
12. Hassan SU, Haddawy P. Measuring international knowledge flows and scholarly impact of scientific research. Scientometrics. 2013;94(1):163-79.
13. Hassan SU, Haddawy P. Analyzing knowledge flows of scientific literature through semantic links: a case study in the field of energy. Scientometrics. 2015;103(1):33-46.

14. CRS Report. Global Research and Development Expenditures: Fact Sheet. 2021. Retrived on 20 March 2022 from https://sgp.fas.org/crs/misc/R44283.pdf
15. Xiaole N. China's research and development spend. Nature. 2015;S8.
16. Raheem S, Hassan S. A bibliometric assessment of scientific productivity and international collaboration of the Islamic World in science and technology (S&T) areas. Scientometrics. 2015;105(2):1059-77.
17. Saeed-Ul H, Sarwar R, Muazzam A. Tapping into intra-and international collaborations of the Organization of Islamic Cooperation states across science and technology disciplines. Science and Public Policy. 2016;43(5):690-701.
18. Haddawy P, Hassan SU, Asghar A, Amin S. A comprehensive examination of the relation of three citation-based journal metrics to expert judgment of journal quality. Journal of Informetrics. 2016;10(1):162-73.
19. Liv L, Bloch CW, Sivertsen G. Options and limitations in measuring the impact of research grants—evidence from Denmark and Norway. Research Evaluation. 2015;24(3):256-70.
20. Waheed H, Hassan SU, Aljohani NR, Wasif M. A bibliometric perspective of learning analytics research landscape. Behaviour and Information Technology. 2016;37(10-11):941-57.
21. Hassan SU, Visvizi A, Waheed H. The 'who'and the 'what'in international migration research: Data-driven analysis of Scopus-indexed scientific literature. Behaviour and Information Technology. 2019;38(9):924-39.
22. Gyorffy B, Herman P, Szabó I. Research funding: past performance is a stronger predictor of future scientific output than reviewer scores. Journal of Informetrics. 2020;14(3):101050.
23. Sabah F, Hassan SU, Muazzam A, Iqbal S, Soroya SH, Sarwar R. Scientific collaboration networks in Pakistan and their impact on institutional research performance: A case study based on Scopus publications. Library Hi-Tech. 2018.
24. Zhu J Hassan SU Mirza HT Xie Q. Measuring recent research performance for Chinese universities using bibliometric methods. Scientometrics. 2014;101(1):429-43.
25. Hassan SU, Haddawy P, Kuinkel P, Degelsegger A, Blasy C. A bibliometric study of research activity in ASEAN related to the EU in FP7 priority areas. Scientometrics. 2012;91(3):1035-51.
26. Hassan SU, Aljohani NR, Idrees N, Sarwar R, Nawaz R, Martínez-Cámara E, Herrera F. Predicting literature's early impact with sentiment analysis in Twitter. Knowledge-Based Systems. 2020;192:105383.
27. Star ZX, Ye FY. h-Efficiency: Measuring input-output performance of research funds. Current Science. 2011;101(1):21-2.
28. Carter B, Graversen EK, Pedersen HS. Competitive research grants and their impact on career performance. Minerva 2014;52(1):77-96.
29. Gallo S, Thompson L, Schmaling K, Glisson S. Risk evaluation in peer review of grant applications. Environment Systems and Decisions. 2018;38(2):216-29.
30. Lutz B, Leydesdorff L, Besselaar PV. A meta-evaluation of scientific research proposals: Different ways of comparing rejected to awarded applications. Journal of Informetrics. 2010;4(3):211-20.
31. Leeuwen V, Thed N, Moed HF. Funding decisions, peer review, and scientific excellence in physical sciences, chemistry, and geosciences. Research Evaluation. 2012;21(3):189-98.
32. Benavente JM, Crespi G, Garone LF, Maffioli A. The impact of national research funds: A regression discontinuity approach to the Chilean FONDECYT. Research Policy. 2012;41(8):1461-75.
33. Brian JA, Lefgren L. The impact of NIH postdoctoral training grants on scientific productivity. Research Policy. 2011;40(6):864-74.
34. Göran M, Danell R. The top eight percent: Development of approved and rejected applicants for a prestigious grant in Sweden. Science and Public Policy. 2006;33(10):702-12.
35. Langfeldt L, Benner M, Sivertsen G, Kristiansen EH, Aksnes DW, Borlaug SB, Pelkonen A. Excellence and growth dynamics: A comparative study of the Matthew effect. Science and Public Policy. 2015;42(5):661-75.
36. Hassan SU, Safder I, Akram A, Kamiran F. A novel machine-learning approach to measuring scientific knowledge flows using citation context analysis. Scientometrics. 2018;116(2):973-96.
37. Ebadi A, Tremblay S, Goutte C, Schiffauerova A. Application of machine learning techniques to assess the trends and alignment of the funded research output. Journal of Informetrics. 2020;14(2):101018.
38. Hassan SU, Akram A, Haddawy P. Identifying important citations using contextual information from full text. In 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL). IEEE. 2017:1-8.
39. Hassan SU, Imran M, Iqbal S, Aljohani NR, Nawaz R. Deep context of citations using machine-learning models in scholarly full-text articles. Scientometrics. 2018;117(3);1645-62.
40. Safder I, Hassan SU, Visvizi A, Noraset T, Nawaz R, Tuarob S. Deep learning-based extraction of algorithmic metadata in full-text scholarly documents. Information processing and management. 2020;57(6):102269.
41. Safder I, Hassan SU. Bibliometric-enhanced information retrieval: A novel deep feature engineering approach for algorithm searching from full-text publications. Scientometrics. 2019;119(1):257-77.
42. Safder I, Sarfraz J, Hassan SU, Ali M, Tuarob S. Detecting target text related to algorithmic efficiency in scholarly big data using recurrent convolutional neural network model. In International conference on Asian digital libraries. Springer, Cham. 2017;30-40.
43. Wolff P, Ríos S, Clavijo D, Graña M, Carrasco M. Methodologically grounded semantic analysis of large volume of Chilean medical literature data applied to the analysis of medical research funding efficiency in Chile. Journal of biomedical semantics. 2020;11(1):1-10.
44. Safder I, Hassan SU. DS4A: Deep search system for algorithms from full-text scholarly big data. In 2018 IEEE international conference on data mining workshops (ICDMW). IEEE. 2018:1308-15.
45. Sami K, Lemola T. Methods for Evaluating the impact of basic research funding: An analysis of recent international evaluation activity. Publications of the Academy of Finland 9.06 (2006).
46. Zaman F, Shardlow M, Hassan SU, Aljohani NR, Nawaz R. HTSS: A novel hybrid text summarization and simplification architecture. Information Processing and Management. 2020;57(6):102351.
47. Hassan SU, Haddawy P. Tapping into Scientific Knowledge Flows via Semantic Links. In ISSI. 2015.
48. Safder I, Batool H, Sarwar R, Zaman F, Aljohani NR, Nawaz R, Hassan SU. Parsing AUC result-figures in machine learning specific scholarly documents for semantically-enriched summarization. Applied Artificial Intelligence. 2021:1-27.
49. Qasim MA, Hassan SU, Aljohani NR, Lytras MD. Human behavior analysis in the production and consumption of scientific knowledge across regions: A case study on publications in Scopus. Library Hi-Tech. 2017.
50. Hassan SU, Iqbal S, Imran M, Aljohani NR, Nawaz R. Mining the context of citations in scientific publications. In International Conference on Asian Digital Libraries. Springer, Cham. 2018:316-22.
51. Chaudhary HAA, Hassan SU, Aljohani NR, Daud A. Quality Classification of Scientific Publications Using Hybrid Summarization Model. In International Conference on Asian Digital Libraries. Springer, Cham. 2018:61-67.
52. Aljohani NR, Fayoumi A, Hassan SU. A Novel Deep Neural Network-Based Approach to Measure Scholarly Research Dissemination Using Citations Network. Applied Sciences, 2021;11(22):10970.
53. Waheed H, Hassan SU, Aljohani NR, Hardman J, Alelyani S, Nawaz R. Predicting academic performance of students from VLE big data using deep learning models. Computers in Human behavior. 2020;104:106189.
54. Hassan SU, Waheed H, Aljohani NR, Ali M, Ventura S, Herrera F. Virtual learning environment to predict withdrawal by leveraging deep learning. International Journal of Intelligent Systems. 2019;34(8):1935-52.
55. Mahmood Z, Safder I, Nawab RMA, Bukhari F, Nawaz R, Alfakeeh AS, Hassan SU. Deep sentiments in roman Urdu text using recurrent convolutional neural network model. Information Processing and Management. 2020;57(4):102233.
56. Safder I, Mahmood Z, Sarwar R, Hassan SU, Zaman F, Nawab RM A, Nawaz R. Sentiment analysis for Urdu online reviews using deep learning models. Expert Systems. 2021;e12751.
57. Safder I, Hassan SU, Aljohani NR. AI cognition in searching for relevant knowledge from scholarly big data, using a multi-layer perceptron and recurrent convolutional neural network model. In Companion Proceedings of the The Web Conference. 2018;:251-8.