An Expertise-based Framework for Research Portfolio Management of Institutions at coarse- and fine-grained levels

Abhirup Nandy[*], Hiran H. Lathabai[**] and Vivek Kumar Singh[***]

[*]*abhirupnandy.online@gmail.com*, [***]*vivek@bhu.ac.in*
0000-0001-8618-0847, 0000-0002-7348-6545
Department of Computer Science, Banaras Hindu University, Varanasi, India.

[**] *hiranhl007@gmail.com*
0000-0002-5633-9842
Amrita CREATE, Amrita Vishwa Vidyapeetham, Amritapuri-690525, Kerala, India.

Abstract: Institutional performance assessment is one of the major challenges for various stakeholders including national and institutional policymakers. Existing popular approaches to performance measurement rely on various factors besides research output, which have been criticized on various grounds. In this work, we present a sciento-text framework to assess the core competency/expertise of an institution at two levels: a broad thematic level, based on WoS subject categories, and a finer thematic level based on indexed keywords. The performance measures namely x_d- index and *x*-index are used for assessment at broad and fine thematic levels, respectively. While national policymakers can make use of x_d- index for the enhancement of national scholarly ecosystem, institutional policymakers and other stakeholders of the institution can make benefit from the wholistic usage of the framework to work for improving its broader expertise diversity as well as enhancing its fine level expertise within suitable disciplines.

Keywords: Expertise Diversity, Expertise Index, Institutional Expertise, Research Portfolio, Research Management.

1. Introduction

The consequences of a recent shift from “trust-based” funding of institutions to “performance-based” assessment is visible in many countries. This change is sometimes facilitated by government and non-government funding agencies globally, who look towards the adoption of comprehensive assessment methods. The major motivation behind adoption of performance-based funding is to ensure the simultaneous determination of – (i) horizontal diversity and pluralism within the system and (ii) vertical differentiation and functional specialization between institutions (Sörlin, 2007). Some examples are– (i) the formation of the Research Excellence Framework (REF) in the UK (Boer et al., 2015), (ii) the allocation of 80 million USD towards a performance-based funding scheme by the Australian government (Maslen, 2019), and (iii) the adoption of the Norwegian model of funding at a national level by Norway, Belgium, Denmark, Finland and Portugal (Sivertsen, 2016). These global activities have pushed institutions to strive for continuous improvement of performance.

To some extent, the rise of major ranking frameworks like the QS, THE, ARWU, and CWTS can be attributed to the above-mentioned shift. These frameworks depend on several factors (which includes research, faculty, funding, etc.) for assessment. However, these frameworks face major criticisms– (i) the ARWU rankings use many irrelevant criteria, and a limited aggregation strategy (Billaut et al., 2010; Jeremic et al., 2011) (ii) the Times (THE) rankings have an anchoring effect (Beck & Morrow, 2010; Bowman & Bastedo, 2011), and (iii) the QS rankings have been commercialized and gives more focus on peer reviews (Anowar et al., 2015). In addition, these rankings lack inclusivity, because many well-performing institutions from the developing countries gets overlooked. These factors forced some countries to go for their own national ranking frameworks, like the National Institutional Ranking Framework (NIRF) in India. However, these frameworks are usually deprived of utilizing the full potential of the bibliometric data, while they also miss out on factors like thematic strengths and areas of expertise. This shortcoming can happen on two levels- (i) a coarse level of overall thematic expertise diversity or broad expertise, and (ii) a fine level of thematic expertise within disciplines.

To overcome these limitations, a network-based framework was introduced by Lathabai et al., (2021a, 2021b). This framework is useful for the analysis of the research portfolio of an institution on a finer level, and uses the keywords used in publications for mapping of publications to fine thematic areas within a discipline. A set of novel indicators, namely the *x*-index and the *x*(*g*)-index, was introduced in this framework. These indicators are inspired by the *h*-index (Hirsch, 2005) and the *g*-index (Egghe, 2006), respectively and are used to determine the core-competency and potential core-competency areas of the institutions. The assessment framework was further developed into a recommendation system framework, where for converting some or all of the potential core competencies of an institution to core competencies, other institutions would be recommended which have corresponding thematic areas as core competency (Lathabai et al., 2022).

On similar grounds, another indicator was also developed for reflecting the expertise and diversity at broad thematic level, which can be computed in similar fashion as that of the *x*-index. This indicator, namely the x_d-index or Expertise Diversity index (Nandy et al., 2023), can be effectively utilized to retrieve coarse level core competency or broader core competency of an institution. This framework uses the WoS subject categories (to represent broad thematic areas or disciplines), which is a curated list of broad thematic areas.

For a comprehensive or wholistic research performance assessment of an institution, we need to analyze both levels of expertise – (i) a broad level core competency to determine the diversity of the research portfolio, and (ii) a fine level core competency within a subject category. The main motivation for this study is the lack of a framework for wholistic research portfolio management that requires determination of expertise at both broad and finer levels. Such a two-level assessment of institutional expertise or research performance will be immensely helpful to policymakers and other stakeholders. The details of such a framework design are discussed next.

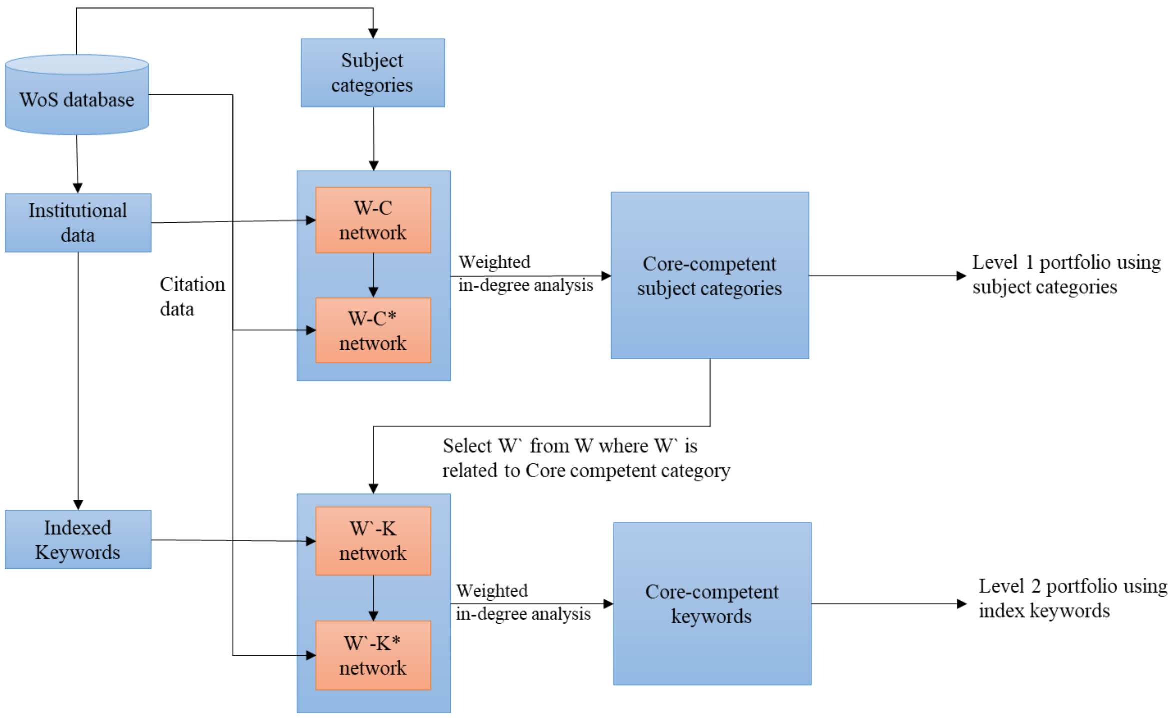
2. Methodology

Network analysis forms the crux of both the broad level as well as find level frameworks. For broad level, the metadata field related to WoS subject category is used and for fine level, the meta data field for keyword is used. Network analysis is mainly used for the formation of work-category affiliation network and work-keyword affiliation network creation and analyses. The schematic diagram of the proposed framework is shown in **Figure 1.** This framework shows how the research portfolio is determined for each institution, at the two different levels. The methodology involves only publication data, which puts more focus on the research output, rather than outside factors that are prone to manipulation.

The proposed methodology uses 4 different fields from the Web of Science data – (i) ‘UT (Unique WOS ID)’, (ii) ‘ID (Keywords Plus)’, (iii) ‘WoS Categories’, and (iv) ‘Z9 (Times Cited, All Databases)’. The data was pre-processed and cleaned based on these fields, before further analysis. The ‘Keywords Plus’ field provides the Index keywords, ‘UT (Unique WOS

ID)' field provides the unique publication IDs, the 'WoS Categories' provides the subject categories, and the 'Times Cited, All Databases' provides the citation information. Using this data, the framework has been divided into two separate sections based on the level of expertise computation– (i) Level 1 – for core-competent WoS subject categories, where the x_d-index is calculated for institutions, and (ii) Level 2 – for core-competent Index keywords, where the *x*-index is calculated within necessary WoS categories.

Figure 1. Framework for determining research portfolio.



2.1. Level 1 – Broad area core competency determination using WoS Subject Categories

The core competent categories for Level 1 are computed based on the concept of the x_d-index. The framework for the x_d-index is based on similar grounds to that of *x*-index (Lathabai et al., 2021a, 2021b), and was adopted on the notion of *h*-index. The indicator x_d-index can be described as –

x_d-index: An institution is supposed to have an x_d-index value of x_d if it has published articles in at least x_d subject categories, and has the strength of at least x_d in those x_d categories. These x_d categories would be considered as the x_d-core competent areas of the institution. A high x_d- index value indicates that the institution's research portfolio is more diverse.

For the computation of the x_d-index, the standard procedure for determination of *h*-index can be done. At first, a W-C (Work-Category) network is created. The W-C network is then transformed into a W-C* network, by "injecting" the citation values through an injection method described by Lathabai et al., (2017). Using the network, the weighted in-degree values of the WoS category nodes are extracted. This will provide the strengths of that institution in different subject categories (broad thematic areas). The subject categories are then sorted and ranked according to the thematic strength values. The x_d-index of the institution is then computed in an *h*-index fashion, by computing the Citation-Rank-Ratio (CRR) and identifying the point where the CRR crosses below 1. In other terms, the x_d is the first occurrence of one of the following cases –

$$x_d = \begin{cases} r, \; if \; CRR = \frac{citation \; at \; poisition \; r}{r} = 1 \\ r-1, \quad if \; CRR = \frac{citation \; at \; poisition \; r}{r} < 1 \end{cases} \qquad (1)$$

So, a WoS category would be considered a core-competency category if CRR ≥ 1 for that category in the institution. Using this approach, all the core competent subject categories C_{core} for an institution are calculated.

2.2. Level 2 – Fine area core competency determination using Index Keywords / Keyword Plus keywords

For a finer level of expertise within a subject category, the *x*-index is used to compute the core-competent keywords within each of the core subject categories. The *x*-index is an indicator which is quite similar to the x_d-index but is based on keywords instead of subject categories. This ensures a finer level of assessment, since keywords are a more specialized set of meta-data for a publication. The *x*-index can be described as –

***x*-index**: An institution is supposed to have an *x*-index value of *x* if it has published papers in at least *x* thematic areas with thematic strengths of at least *x*. Here the thematic strengths are computed as total citation scores or altmetric scores received for those areas. These *x* areas that form the *x*-core can be treated as the core competency areas of the institution.

Here, each of the core-competent categories $c \in C_{core}$ is taken iteratively, and the list of core-competent keywords within c is calculated. This is done by extracting a subnetwork WC_c from the WC network, where the list of publications W' is restricted to only those which have category c in their publication metadata while taking each $c \in C_{core}$. Using this W`, we create a W`K or Work-Keyword network. Using the W`K network, a similar approach was used as described in $x_d$-index to compute *x*-index within that category. W`-K network is converted to W`-K* network using injection approach. The keywords are then ranked, and a ratio of the in-degree value to the ranks is obtained for each keyword. The list of core-competent keywords K_{core} is then obtained, where any keyword $k \in K_{core}$ would have the CRR ratio ≥ 1. This gives us a list of core-competent keywords K^c_{core}, for each of the category $c \in C_{core}$. A bridged version of the portfolio for "University of Madras", which has a x_d-index of 89, is shown in **Figure 2.**

The two-level list retrieved for each institution is then used to rank institutions and subject categories. We can use the x_d-index to rank institutions based on core-competent categories, and further rank the categories with the *x*-index computed using core-competent keywords.

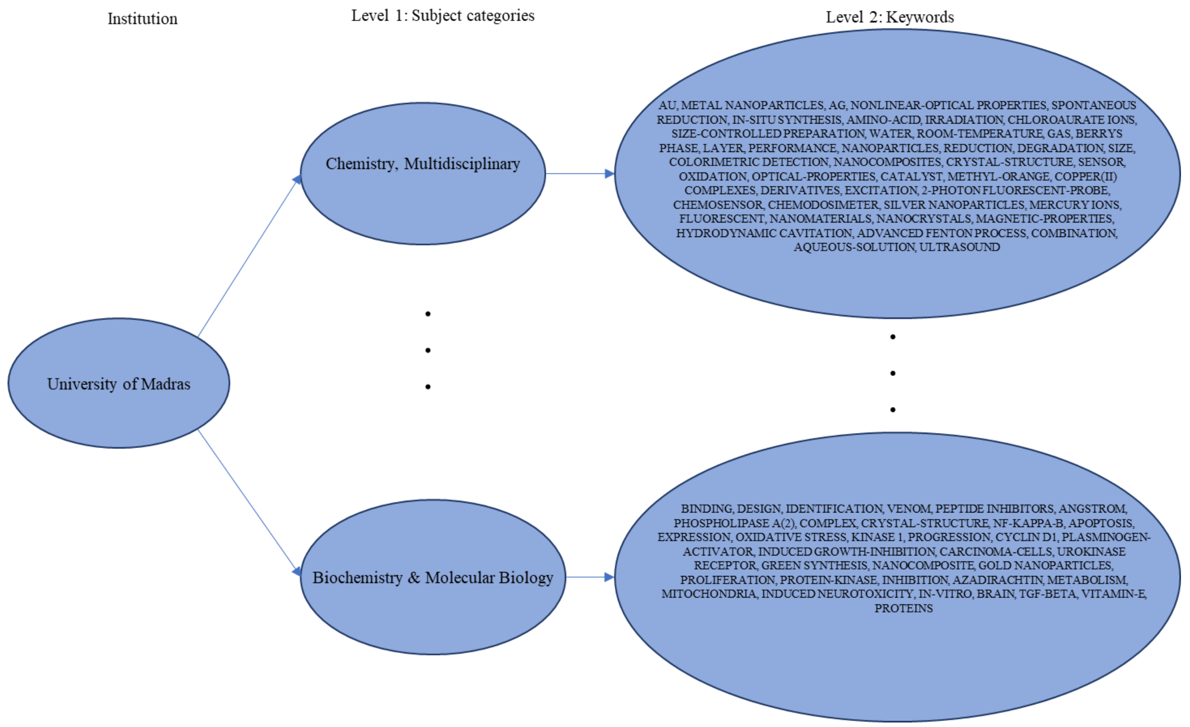
3. Data

The article meta-data was collected from a list of 136 Indian Institutions from WoS, which were ordered based on their number of publications. This list excluded all possible observations of institutional systems comprising of multiple institutions like the IIT system and included the individual institutions only. A total of 467,550 articles were fetched and further used for the study. Although the study represented data from 2011 to 2020 only, the framework itself is easily capable of being effective for a larger span of data if needed. Similarly, this exercise can be done for data at different intervals to determine the expertise of institutions at various points of time. **Table 1** provides more insights about the data. For the data about Indian institutions, it was found that publications span across 250 WoS subject categories, and there are 292,267 Keyword Plus (or Index) keywords from the whole dataset.

Table 1. Description of the WoS data used.

No. of institutions used in the study	Total no. of articles retrieved	Total no. of WoS subject categories	Total no. of WoS Index Keywords
136	467,550	250	292,267

Figure 2. The two-level portfolio of an example institution - University of Madras (the index values are not included in the figure)



4. Results

From the whole data for 136 Indian institutions, we have calculated the x_d-index and x-index for the full data. The analysis shows that "University of Delhi" has the highest x_d-index of 156, followed by "Banaras Hindu University BHU" with an x_d-index of 140. This means University of Delhi has publications in 156 WoS subject categories, where it has at least 156 citations in each. Similarly, BHU has publications in 140 subject areas with at least 140 citations in each. The lowest x_d-index value was for "Inter University Accelerator Centre", with 36 subject areas with at least 36 citations. This shows that the institutions with high x_d-index values have a diverse research portfolio, while institutions with relatively lower x_d-index values might have more focused research areas. The full list of 136 institutions with their x_d-index is shown in **Figure 3**. The x_d-index values are a reflection of disciplinary diversity/ expertise of these institutions.

Figure 3. The x_d-index values for the 136 institutions.

Sr. No.	Institution	Total Publications	x_d-index
1	ACADEMY OF SCIENTIFIC INNOVATIVE RESEARCH ACSIR	9,972	98
2	ALAGAPPA UNIVERSITY	2,347	81
3	ALIGARH MUSLIM UNIVERSITY	6,724	119
4	ALL INDIA INSTITUTE OF MEDICAL SCIENCES AIIMS NEW DELHI	8,959	103
5	AMITY UNIVERSITY NOIDA	2,405	100
6	AMRITA VISHWA VIDYAPEETHAM	2,856	101
7	ANDHRA UNIVERSITY	2,093	81
8	ANNA UNIVERSITY	9,960	110
9	ANNAMALAI UNIVERSITY	3,976	95
10	BANARAS HINDU UNIVERSITY BHU	11,765	140
11	BHARATHIAR UNIVERSITY	4,262	97
12	BHARATHIDASAN UNIVERSITY	3,139	89
13	BIRLA INSTITUTE OF TECHNOLOGY MESRA	2,276	87
14	BIRLA INSTITUTE OF TECHNOLOGY SCIENCE PILANI BITS PILANI	4,616	109
15	BOSE INSTITUTE	2,016	67
16	CHRISTIAN MEDICAL COLLEGE HOSPITAL CMCH VELLORE	2,718	68
17	COCHIN UNIVERSITY SCIENCE TECHNOLOGY	2,386	89
18	CSIR CENTRAL DRUG RESEARCH INSTITUTE CDRI	3,068	69
19	CSIR CENTRAL ELECTROCHEMICAL RESEARCH INSTITUTE CECRI	2,244	51
20	CSIR CENTRAL FOOD TECHNOLOGICAL RESEARCH INSTITUTE CFTRI	1,939	52
21	CSIR CENTRAL GLASS CERAMIC RESEARCH INSTITUTE CGCRI	1,664	50
22	CSIR CENTRAL LEATHER RESEARCH INSTITUTE CLRI	2,023	64
23	CSIR CENTRAL SALT MARINE CHEMICAL RESEARCH INSTITUTE CSMCRI	1,991	55
24	CSIR CENTRE FOR CELLULAR MOLECULAR BIOLOGY CCMB	1,000	57
25	CSIR INDIAN INSTITUTE OF CHEMICAL BIOLOGY IICB	1,961	70
26	CSIR INDIAN INSTITUTE OF CHEMICAL TECHNOLOGY IICT	6,153	72
27	CSIR INSTITUTE OF GENOMICS INTEGRATIVE BIOLOGY IGIB	1,408	69
28	CSIR NATIONAL CHEMICAL LABORATORY NCL	4,930	69
29	CSIR NATIONAL INSTITUTE INTERDISCIPLINARY SCIENCE TECHNOLOGY NIIST	2,059	59
30	CSIR NATIONAL INSTITUTE OF OCEANOGRAPHY NIO	1,995	54
31	CSIR NATIONAL PHYSICAL LABORATORY NPL	3,476	63
32	DELHI TECHNOLOGICAL UNIVERSITY	1,973	92
33	DR B R AMBEDKAR NATIONAL INSTITUTE OF TECHNOLOGY JALANDHAR	1,640	72
34	GAUHATI UNIVERSITY	1,814	76
35	GOVT MED COLL	1,187	58
36	GURU NANAK DEV UNIVERSITY	3,371	91
37	ICAR INDIAN AGRICULTURAL RESEARCH INSTITUTE	5,123	65
38	ICAR INDIAN VETERINARY RESEARCH INSTITUTE	2,716	54
39	ICAR NATIONAL DAIRY RESEARCH INSTITUTE	2,260	45
40	INDIAN ASSOCIATION FOR THE CULTIVATION OF SCIENCE IACS JADAVPUR	4,526	53
41	INDIAN INSTITUTE OF ENGINEERING SCIENCE TECHNOLOGY SHIBPUR IIEST	3,342	82
42	INDIAN INSTITUTE OF SCIENCE EDUCATION RESEARCH IISER BHOPAL	1,830	59
43	INDIAN INSTITUTE OF SCIENCE EDUCATION RESEARCH IISER KOLKATA	2,915	80
44	INDIAN INSTITUTE OF SCIENCE EDUCATION RESEARCH IISER MOHALI	1,771	60
45	INDIAN INSTITUTE OF SCIENCE IISC BANGALORE	18,098	132
46	INDIAN INSTITUTE OF TECHNOLOGY IIT BHU VARANASI	5,121	99
47	INDIAN INSTITUTE OF TECHNOLOGY IIT BOMBAY	13,821	122
48	INDIAN INSTITUTE OF TECHNOLOGY IIT DELHI	12,938	130
49	INDIAN INSTITUTE OF TECHNOLOGY IIT GANDHINAGAR	1,673	73
50	INDIAN INSTITUTE OF TECHNOLOGY IIT GUWAHATI	8,582	114
51	INDIAN INSTITUTE OF TECHNOLOGY IIT HYDERABAD	3,186	89
52	INDIAN INSTITUTE OF TECHNOLOGY IIT INDORE	3,169	84
53	INDIAN INSTITUTE OF TECHNOLOGY IIT KANPUR	9,882	116
54	INDIAN INSTITUTE OF TECHNOLOGY IIT KHARAGPUR	15,498	137
55	INDIAN INSTITUTE OF TECHNOLOGY IIT MADRAS	14,132	126
56	INDIAN INSTITUTE OF TECHNOLOGY IIT PATNA	1,818	72
57	INDIAN INSTITUTE OF TECHNOLOGY IIT ROORKEE	10,548	125
58	INDIAN INSTITUTE OF TECHNOLOGY IIT ROPAR	1,718	76
59	INDIAN INSTITUTE OF TECHNOLOGY INDIAN SCHOOL OF MINES DHANBAD	6,040	99
60	INDIAN SPACE RESEARCH ORGANISATION ISRO	4,041	72
61	INDIAN STATISTICAL INSTITUTE	3,845	95
62	INDIRA GANDHI CENTRE FOR ATOMIC RESEARCH IGCAR	3,831	70
63	INSTITUTE OF CHEMICAL TECHNOLOGY MUMBAI	3,780	66
64	INTER UNIVERSITY ACCELERATOR CENTRE	1,691	36

65	JADAVPUR UNIVERSITY	9,427	115
66	JAMIA HAMDARD UNIVERSITY	2,904	82
67	JAMIA MILLIA ISLAMIA	4,155	110
68	JAWAHARLAL INSTITUTE OF POSTGRADUATE MEDICAL EDUCATION RESEARCH	1,409	51
69	JAWAHARLAL NEHRU CENTER FOR ADVANCED SCIENTIFIC RESEARCH JNCASR	2,992	66
70	JAWAHARLAL NEHRU TECHNOLOGICAL UNIVERSITY HYDERABAD	1,758	72
71	JAWAHARLAL NEHRU UNIVERSITY NEW DELHI	4,927	111
72	KALINGA INSTITUTE OF INDUSTRIAL TECHNOLOGY KIIT	1,000	88
73	KALYANI UNIVERSITY	2,135	80
74	KASTURBA MEDICAL COLLEGE MANIPAL	1,411	71
75	KURUKSHETRA UNIVERSITY	1,665	78
76	L V PRASAD EYE INSTITUTE	1,541	39
77	LOVELY PROFESSIONAL UNIVERSITY	1,658	78
78	LUCKNOW UNIVERSITY	1,979	80
79	MADURAI KAMARAJ UNIVERSITY	2,273	73
80	MAHARAJA SAYAJIRAO UNIVERSITY BARODA	2,219	88
81	MAHARSHI DAYANAND UNIVERSITY	1,585	73
82	MAHATMA GANDHI UNIVERSITY KERALA	1,568	69
83	MALAVIYA NATIONAL INSTITUTE OF TECHNOLOGY JAIPUR	2,169	80
84	MANIPAL ACADEMY OF HIGHER EDUCATION MAHE	5,955	125
85	MAULANA AZAD MEDICAL COLLEGE	1,198	45
86	MOTILAL NEHRU NATIONAL INSTITUTE OF TECHNOLOGY	1,966	74
87	NATIONAL CENTRE FOR BIOLOGICAL SCIENCES NCBS	1,512	64
88	NATIONAL INSTITUTE OF MENTAL HEALTH NEUROSCIENCES INDIA	2,498	61
89	NATIONAL INSTITUTE OF PHARMACEUTICAL EDUCATION RESEARCH MOHALI	1,444	57
90	NATIONAL INSTITUTE OF TECHNOLOGY CALICUT	1,838	80
91	NATIONAL INSTITUTE OF TECHNOLOGY DURGAPUR	2,401	82
92	NATIONAL INSTITUTE OF TECHNOLOGY KARNATAKA	2,833	84
93	NATIONAL INSTITUTE OF TECHNOLOGY KURUKSHETRA	1,674	71
94	NATIONAL INSTITUTE OF TECHNOLOGY ROURKELA	4,938	107
95	NATIONAL INSTITUTE OF TECHNOLOGY SILCHAR	1,657	66
96	NATIONAL INSTITUTE OF TECHNOLOGY TIRUCHIRAPPALLI	4,229	82
97	NATIONAL INSTITUTE OF TECHNOLOGY WARANGAL	2,204	71
98	PHYSICAL RESEARCH LABORATORY INDIA	2,137	43
99	PONDICHERRY UNIVERSITY	3,171	96
100	PGIMER CHANDIGARH	6,441	85
101	PUNJAB AGRICULTURAL UNIVERSITY	2,464	56
102	PUNJABI UNIVERSITY	2,273	96
103	RAJA RAMANNA CENTRE FOR ADVANCED TECHNOLOGY	1,803	47
104	RASHTRASANT TUKADOJI MAHARAJ NAGPUR UNIVERSITY	1,535	66
105	SANJAY GANDHI POSTGRADUATE INSTITUTE OF MEDICAL SCIENCES	2,256	66
106	SARDAR VALLABHBHAI NATIONAL INSTITUTE OF TECHNOLOGY	2,140	80
107	SATHYABAMA INSTITUTE OF SCIENCE TECHNOLOGY	1,543	74
108	SAVITRIBAI PHULE PUNE UNIVERSITY	4,246	98
109	SETH GORDHANDAS SUNDERDAS MEDICAL COLLEGE KING EDWARD MEMORIAL HOSPITAL	1,059	52
110	SHANMUGHA ARTS SCIENCE TECHNOLOGY RESEARCH ACADEMY SASTRA	3,354	95
111	SHIVAJI UNIVERSITY	2,467	70
112	SIKSHA O ANUSANDHAN UNIVERSITY	1,946	80
113	SN BOSE NATIONAL CENTRE FOR BASIC SCIENCE SNBNCBS	1,954	49
114	SREE CHITRA TIRUNAL INSTITUTE FOR MEDICAL SCIENCES TECHNOLOGY SCTIMST	1,323	64
115	SRI VENKATESWARA UNIVERSITY	2,361	75
116	SRM INSTITUTE OF SCIENCE TECHNOLOGY CHENNAI	4,640	106
117	SSN COLLEGE OF ENGINEERING	1,613	59
118	ST JOHN S NATIONAL ACADEMY OF HEALTH SCIENCES	1,209	60
119	TATA MEMORIAL CENTRE TMC	2,353	63
120	TATA MEMORIAL HOSPITAL	2,006	59
121	TEZPUR UNIVERSITY	2,742	90
122	THAPAR INSTITUTE OF ENGINEERING TECHNOLOGY	5,141	98
123	UGC DAE CONSORTIUM FOR SCIENTIFIC RESEARCH	2,245	43
124	UNIVERSITY COLLEGE OF MEDICAL SCIENCES	1,033	54
125	UNIVERSITY OF ALLAHABAD	2,335	94
126	UNIVERSITY OF BURDWAN	2,161	84
127	UNIVERSITY OF CALCUTTA	7,405	123
128	UNIVERSITY OF DELHI	12,994	156
129	UNIVERSITY OF HYDERABAD	5,361	102
130	UNIVERSITY OF JAMMU	1,644	66
131	UNIVERSITY OF KASHMIR	1,879	89
132	UNIVERSITY OF MADRAS	3,324	89
133	UNIVERSITY OF MYSORE	2,253	76
134	UNIVERSITY OF RAJASTHAN	1,712	75
135	VELLORE INSTITUTE OF TECHNOLOGY	8,153	118
136	VISVESVARAYA NATIONAL INSTITUTE OF TECHNOLOGY NAGPUR	2,129	80

The x_d-index values are compared with h-index, g-index and the Shannon's Entropy. Shannon's Entropy is an indicator used to verify the standard diversity measure. The SRCC value of the x_d-index based rankings with that h-index and g-index are 0.6013 and 0.4437 respectively, suggesting that x_d-index is different from these indicators. The SRCC value of x_d-index with Shannon's Entropy value is 0.8648, indicating a high correlation. The h-index and g-indices, on the other hand, have SRCC of 0.2791 and 0.1932 with Shannon's Entropy, which tells that they cannot be effectively used to measure the diversity of the portfolio, while our proposed framework is more capable of demonstrating the diversity.

While our study incorporates the use of both the x-index and the x_d-index, the finer thematic areas extracted using x-index provides more information like specificities of the research within a broad area of expertise of an institution. For example, the x-index of the subject category "Chemistry, multidisciplinary" for "University of Madras" is 45, which means there are 45 core competent keywords within the category, which have at least 45 citations. This framework thus showcases both the diversity as well as the quality of the research portfolio of an institution. Both of these indices are necessary for the framework, since they provide information at two different levels. The SRCC between the overall x-index and the x_d-index for the institutions is 0.6946, which shows that they are positively correlated, and should be simultaneously used within the framework.

5. Discussion

A comprehensive portfolio is a vital resource for institutional as well as national level policymakers, researchers, and other academicians. The proposed methodology focusses on the core-competent research categories, and further into the core-competent keywords within the research areas for each of the 136 institutions. A higher value of x_d-index would reflect that the institution has good quality research in a higher number of WoS subject categories. Although this index is quite similar to the h-index, the latter only demonstrates the overall quality and quantity of research for an institution and fails to bring out how diverse the research area of the institution is.

The use of WoS subject categories as a level 1 portfolio has many benefits. At this level, the portfolio is formed using x_d-index, which uses the WoS subject categories for performance assessment. The WoS category list for each publication is a subset of the 254 subject categories in the WoS database. This is a curated list and is selected based on the publication source details of the publication (Singh et al., 2020). The use of broad subject categories also helps in studying the institutional level diversity. This can be used to make decisions like expansion of more research areas within an institution on a broader scale (for example, establishing a new department), or the policymaker choosing an institution for further collaboration, based on the broad subject categories in which it excels at.

Along with the broad level assessment, a second level of the portfolio is also presented. This is to determine the finer level thematic areas of research within the core subject categories, using the x-index. The x-index, when proposed, used an NLP module since the work was with Author-provided keywords, which is prone to redundancy and errors of various kinds (Lathabai et al., 2021b). Rather, we propose the use of Index keywords ("Keywords Plus" field of the metadata), which is extracted using various algorithms and is less prone to the previous issues. This ensures a refined set of keywords for computing the finer-level core competency of the institution. This level of the portfolio can be used to determine which specific themes the institution is working on, within the core subject categories. This can be used in applications like selecting an individual/group within a core-competent department of an institution for collaboration, who has been working on the core-competent keyword.

This two-level portfolio can be used by institutional level policy-makers to keep a track of the core-competent broad level subject categories as well as further finer level keywords which the institution excels at. This research portfolio can be used to induce collaboration possibilities between institutions which lack core-competency in a certain subject area, with an institution that has a core competency in the same. This can also be used to put more focus on keywords which are not core-competent within a core-competent subject category, and thus further enhance the quality of research in the specific category within an institution.

National level policymakers can also effectively use the research portfolio to further enhance the overall research diversity of an institution and the country as well. Such policy makers may take decisions like –

(i) Develop policies for establishing novel research collaboration between institutions with similar core-competency at either one or at both levels of expertise. Such collaborations may be among Academic institutions themselves (A2A), with the government (A2G), or even with the industry (A2I).
(ii) Develop policies for further growth of international collaborations based on the two levels of expertise.

Although the proposed indicator can be used to compute the diversity of an institution at two different levels, the methodology has been tried on WoS database only. The robustness of the framework can be affirmed if a different database is used, like the Scopus database (which contains Subject Areas for level 1, and author keywords for level 2), or the Dimensions database (which contains the FOR field for level 1, and concepts for level 2). This extension of the current work would be reserved for further study.

6. Conclusion

In this study, we have proposed a framework for a research portfolio of an institution. This research portfolio consists of two levels – (i) a broad level thematic area classification to determine the core competent subject categories in which an institution excels, using an Expertise index x_d-index, and (ii) a finer level thematic area classification, to determine the core competent keywords within the core competent categories. This two-level research portfolio may benefit institutional as well as national-level policy makers. Institutional policymakers can use the portfolio to showcase their core competent research areas and keywords to other institutions for further possibilities of collaborations. National level policymakers can use the institutional portfolios to define policies based on institutions with similar portfolios, or propose international collaborations. This framework can be easily used to enhance the scholarly ecosystem of an institution, and present an institution’s research interests at two different levels.

Open science practices

This work used research publication data for 136 Indian institutions for the period 2011-20 from the Web of Science database. We will be happy to share the publication DOIs on request.

The analysis and framework designed mainly utilized computer programs written in Python and would be shared on request.

Author contributions

The first author downloaded the data, carried out experimental work and participated in writing of the paper. The second author proposed the idea of expertise-based indices and participated in writing and review. The third author conceptualized the work and guided the experimental work and participated in writing and review of the paper.

Competing interests

The authors declare that manuscript complies with ethical standards of the conference and there is no conflict of interests whatsoever.

Funding information

This work is partly supported by extramural research grant no.: MTR/2020/000625 from Science and Engineering Research Board (SERB), India, and by HPE Aruba Centre for Research in Information Systems at BHU (No.: M-22-69 of BHU).

References

Anowar, F., Helal, M. A., Afroj, S., Sultana, S., Sarker, F., & Mamun, K. A. (2015). A critical review on world university ranking in terms of top four ranking systems. *Lecture Notes in Electrical Engineering*, 312, 559–566. https://doi.org/10.1007/978-3-319-06764-3_72

Beck, S., & Morrow, A. (2010). Canada's universities make the grade globally. The Globe And Mail.

Billaut, J. C., Bouyssou, D., & Vincke, P. (2010). Should you believe in the Shanghai ranking? *Scientometrics*, 84(1), 237–263. https://doi.org/10.1007/s11192-009-0115-x

Boer, H. F. de, Jongbloed, B. W. A., Benneworth, P. S., Cremonini, L., Kolster, R., Kottmann, A., Lemmens-Krug, K., & Vossensteyn, J. J. (2015). Performance-based funding and performance agreements in fourteen higher education systems. Center for Higher Education Policy Studies (CHEPS). https://research.utwente.nl/en/publications/performance-based-funding-and-performance-agreements-in-fourteen-

Bowman, N. A., & Bastedo, M. N. (2011). Anchoring effects in world university rankings: Exploring biases in reputation scores. *Higher Education*, 61(4), 431–444. https://doi.org/10.1007/s10734-010-9339-1

Egghe, L. (2006). An improvement of the h-index: The g-index. *ISSI Newsletter*.

Hirsch, J. E. (2005). An index to quantify an individual's scientific research output. *PNAS*, 102(46), 16569–16572. https://doi.org/10.1073/pnas.0507655102

Jeremic, V., Bulajic, M., Martic, M., & Radojicic, Z. (2011). A fresh approach to evaluating the academic ranking of world universities. *Scientometrics*, 87(3), 587–596. https://doi.org/10.1007/s11192-011-0361-6

Lathabai, H. H., Nandy, A., & Singh, V. K. (2021a). Expertise-based institutional collaboration recommendation in different thematic areas. CEUR Workshop Proceedings, 2847.

Lathabai, H. H., Nandy, A., & Singh, V. K. (2021b). x-index: Identifying core competency and thematic research strengths of institutions using an NLP and network based ranking framework. *Scientometrics*, 126(12), 9557–9583. https://doi.org/10.1007/S11192-021-04188-3/TABLES/5

Lathabai, H. H., Nandy, A., & Singh, V. K. (2022). Institutional collaboration recommendation: An expertise-based framework using NLP and network analysis. *Expert Systems with Applications*, 209, 118317. https://doi.org/10.1016/j.eswa.2022.118317

Lathabai, H. H., Prabhakaran, T., & Changat, M. (2017). Contextual productivity assessment of authors and journals: a network scientometric approach. *Scientometrics*, 110(2), 711–737. https://doi.org/10.1007/S11192-016-2202-0/TABLES/7

Maslen, G. (2019, August 24). New performance-based funding system for universities. https://www.universityworldnews.com/post.php?story=20190822085127986

Nandy, A., Lathabai, H. H., & Singh, V. K. (2023). x_d-index: An overall scholarly expertise index for the research portfolio management of institutions. *Accepted to appear in Proceedings of ISSI2023*.

Singh, P., Piryani, R., Singh, V. K., & Pinto, D. (2020). Revisiting subject classification in academic databases: A comparison of the classification accuracy of Web of Science, Scopus & Dimensions. *J. Intell. Fuzzy Syst.*, 39(2), 2471–2476. https://doi.org/10.3233/JIFS-179906

Sivertsen, G. (2016). Publication-based funding: The norwegian model. In Research Assessment in the Humanities: Towards Criteria and Procedures. https://doi.org/10.1007/978-3-319-29016-4_7

Sörlin, S. (2007). Funding diversity: Performance-based funding regimes as drivers of differentiation in higher education systems. *Higher Education Policy*, 20(4), 413–440. https://doi.org/10.1057/palgrave.hep.8300165