COLLNET Journal of Scientometrics and Information Management
ISSN : 0973 – 7766 (Print) 2168 – 930X (Online)
Vol. 19(2) December 2025, pp. 197–214
DOI : 10.47974/CJSIM-2025-01003

80/20 law in Information Production Processes (IPPs)

Hiran Haridas Lathabai
Solanki Gupta
Rakhi Mohan
Vivek Kumar Singh

Hiran Haridas Lathabai *
Amrita Center for Research in Analytics, Education and Training
Amrita Vishwa Vidyapeetham
Amritapuri Campus
Kollam 690525
Kerala
India

Solanki Gupta †
School of Computer Science Engineering and Technology
Bennett University
Greater Noida 201310
Uttar Pradesh
India

Rakhi Mohan §
Department of Statistics
Government College
Attingal
Thiruvananthapuram 695101
Kerala
India

Vivek Kumar Singh ‡
Department of Computer Science
University of Delhi
Delhi 110007
India

and

NITI AYOG
Government of India

ORCID-ID
* 0000-0002-5633-9842

E-mail
* hiranhl007@gmail.com; hiranhl@am.amrita.edu
† solankigupta2@gmail.com
§ rakhiammalu@gmail.com
‡ vivekks12@gmail.com

* (Corresponding Author)

Information Production Processes (IPPs) like source-item systems possess several properties. Skewness is one property that makes indicators such as h-type indicators relevant for determining key sources that produced the majority of items. Due to this skewness, the 80/20 rule (in which the top 80 percent of items may be produced by the top 20 percent of the sources) also is at play in IPPs. In this work, we investigate whether any of the major existing h-type indicators are capable of reflecting the 80/20 rule effectively. Also, can considering 20% of T (total publications) or 0.2T as an indicator, provide a better alternative? Or is there a room for new indicator or perhaps a new law? These questions are answered in this work. The implications and potential uses of the most suitable base indicator for scientometric applications, especially for science, technology and innovation (STI) management and for applications outside the realm of scientometrics are discussed.

Keywords: *80/20 law, Information Production Processes (IPPs), Skewness, h-type indicators, Ideal indicator, y-index.*

1. Introduction

Many real-world systems are characterized by production processes that link causes/sources (of production) to outcomes. Pareto [1] came up with the 80/20 rule or Pareto principle according to which, 20% of the causes/sources are responsible for 80% of outcomes. For instance, roughly 20% of plants in a garden produce almost 80% of fruits. Similarly, roughly 80% of land might be owned by 20% of people. Therefore, this rule is found to be relevant in so many systems where there is inherent skewness in the distribution of outcomes with respect to sources/ causes, physical as well as information systems.

The Information Production Process (IPP) is a fundamental concept focused solely on information generation, its dissemination, and utilization within the scholarly ecosystem. IPP provides a generalized mathematical framework that describes the whole Information Production Process using source-item terminology. This framework consists of a triple (S, I, F) including a set of sources (S), a set of items (I), and a mapping function F: S → I. This function explains how various sources contribute to the production of items within a system [2]. Examples of IPPs include but not limited to authors and their articles (author-paper systems), journals and their articles (journal-paper systems), articles and their received citations (paper-citation systems), books and their borrowings, words and their occurrence, websites and hyperlinks, etc., [3]. Interestingly, skewness is inherent in these IPPs as small fraction of sources produce major fraction of items, indicating the possibility of 80/20 law at play.

1.1 *80/20 law in IPPs*

Burrel's work [4] that related mean borrowings and proportion of books required to account for 80% of borrowings is one of the earliest explorations of the 80/20 law in IPPs. Egghe [5] shown that for IPPs with Lotka-type distributions (i.e., conforms with Lotka's law) like journal-article, author-article and word-occurrence IPPs, 80/20 effect is much stronger. Egghe [6] also derived and tested the relationship between Pratt's class concentration measure and the fraction of sources (that produces large fraction of items as in 80/20 rule) in case of Lotka's law. Further, the relationship between Price's law and Lotka's law was also derived using the theory of 80/20 rule by Egghe [7]. While journal-article, author-article and word-occurrence IPPs are well-studied in the context of 80/20 law, paper-citation systems are underexplored. This gap is attempted to be filled in this work.

1.2 *Paper-citation IPP and the 80/20 law*

In the paper-citation informetric systems, research papers serve as the primary sources of scientific knowledge, while citations reflect their recognition and influence within the scholarly community. Hence, this relationship exemplifies the dynamics of knowledge dissemination and academic influence.

The skewness inherent in the paper-citation systems is the key for determining the important sources (papers) that are responsible for gathering a great chunk of items (citations), making various indicators for the assessment of productivity and impact or both relevant. Hirsch considered this paper-citation relation to demonstrate the *h*-index [8]. The *h*-index of a scientist/inventor indicates that he/she has at least *h* papers each of which received at least *h* citations (or attention score or any other suitable score). *h*-index serves as a quantitative evaluation measure and provides an indication of the core among the publications having a high 'density' of impact, popularly known as h-core. Studies like [9] explored its useful properties. Several other indicators were introduced with the same/similar underlying computation mechanism as that of the h-index to address the inherent issues in the h-index. These are popularly known as h-type indicators. Indices like $h(2)$-index [10], g-index [11], [12], f and t indices [13], and Ψ-index [14] are some of the prominent h-type indicators.

h-index and its above-mentioned variants are supposed to provide the set of most important sources/papers that gathered a major share of citations (in other words, the set of most impactful sources), as these indicators are dependent on the (skewness of) distribution of citations (items) at varying extent [15]. Thus, in a way, the 80/20 rule is possibly at play in IPPs like article-citation systems as in the case of other IPPs observed by [4], [5]. In other words, the original h-index and its major variants (mentioned in the previous paragraph) are supposed to reflect or get as close as possible to reflect the 80/20 law at work. Of course, it is difficult to find perfect 80/20 rule-abiding real-world IPPs including paper-citation informetric systems. However, in spite of the existence of so many indicators that are direct variants of the h-index, the fact that the ability of such indices to reflect the 80/20 rule still remains unexplored is intriguing. Predominant research focuses on the limitations [16], [17], [18], [19] of the h-index and its direct variants such as the inability to reflect the (i) effect of multiple authorship [20], [21], (ii) effect of fields [22] to which publications (sources) belong, etc., that led to development of some indicators like generalized h and g indices [23], highly cited papers or HCP [24], [25], $h\alpha$ [26], etc., that are claimed be free from one or more of these limitations or offer some other improvements, can be a major reason. While it is possible to develop indicators that reflect the effect of multiple authorship, field bias effect (of citations), etc., by suitably modifying the h-type indicator, if the original/base h-type indicator is not properly reflecting the 80/20 rule at least in a reasonable majority of real-world profiles (reflecting 80/20 rule in all the profiles is not possible), usage of modified indicators for assessment might also be defaulted. Thus, it is vital to probe the ability of h-type indicators to reflect the 80/20 law. If law-indicator compatibility between h-type indicators and 80/20 law cannot be affirmed by that investigation, other indicators might be needed or perhaps another law is at play instead of 80/20 law.

1.3 *Possibilities of other laws discussed in earlier works*

Also, the 80/20 law faced many criticisms too. One of the major criticisms is there is nothing special about the 80/20 law and real-world systems obey other laws too, like 60/40, 70/30, 90/10, etc. For instance, Burrel's and Egghe's early works on 80/20 law in IPPs (other than paper-citation systems) also remarked that 80/20 is one of the possible cases [4], [5]. Urquhart [27] observed that 10% of the journals accounted for 80% of the loan/grant allowed to purchase journals. Upon a study on citation concentration around journals, Garfield [28] noticed that a relatively small core of journals garnered 90% of the citations and stated that such a small core of multidisciplinary journals can satisfy the majority of the scientific information needs. Bensman [29] revealed the mutual complementing nature of Urquhart's law and Garfield's law, thereby resolving a persistent controversy. Rousseau et al. [3], [30] discussed a generalized rule viz. 100y/100x rule (of which 80/20 is one of the cases). They proposed Lorenz curve and Gini index to determine the unbalanced situation induced by 100y/100x rule in case of author-article IPPs. However, these works have not explored 80/20 law in paper-citation IPPs, especially the relationship between various *h*-type indicators and their compliance with the 80/20 law.

Thus, upon our investigation of the law-indicator compliance of *h*-type indicators with 80/20 law, if existing indicators are not found substantially compatible with the 80/20 law, perhaps another law might be suitable for IPPs like paper-citation systems. However, the possibilities of new law or better indicator can be explored only if compatibility of 80/20 law and existing indicators are investigated. This served as our motivation for the current study. Specifically, the following stated investigation points will be beneficial for scientometricians who are into indicator research.

2. Objectives

1. To investigate which among the existing *h*-type indicators like *h(2), h, f, t,* and *g* is most capable of reflecting 80/20 law in IPPs (especially paper-citations systems).
2. To investigate whether the indicator obtained by considering 20% of T (i.e., $0.2T$ as an indicator) reflects the 80/20 law in a better way than the existing indicators.
3. To investigate whether the best indicator found after completion of the first two objectives is most suitable for reflecting the 80/20 law?

If the answers obtained for both questions are not positive, the need for a new indicator or a new law can be confirmed. To investigate the above objectives, we selected *h*-index and its famous variants like *h(2), h, f, t,* and *g*. These indicators are selected as these indicators are computed in more or less the same fashion as that of the *h*-index (satisfying different criteria). Though Ψ -index is also such an indicator, as it is designed to reflect 'offsetability' [14], [31] and is more likely to be found at the other side of the spectrum, the deviation of Ψ -index from 0.2 T and citations up to Ψ (i.e., $C(\Psi)$) from 0.8 C_T might be higher in most of the profiles. Thus, the possibility of limitation of Ψ -index to reflect the 80/20 law is also higher. The details of the investigation on other indicators are described next.

3. Data and Methodology

To investigate the above two objectives, a real-world dataset of scholar profiles is required. We use the dataset used by Lathabai [14] for this purpose. This dataset comprises profiles of 50 eminent scholars in the field of scientometrics, who served as distinguished reviewer board members of the journal 'Scientometrics' as of 14 July 2020.

Objective 1: For investigating whether the existing *h*-type indicators like *h(2), h, f, t,* and *g* are capable of reflecting 80/20 law in IPPs (especially paper-citations systems), the following steps are required for each profile.

1. All the source-item pairs corresponding to these indicators are found out. That means the indicator-citation pairs of all the indicators such as *h(2) & C(h(2)), h and C(h), f and C(f), t and C(t), g and C(g)* are computed.
2. Let (I, C_I) be the ideal 80/20 rule reflecting the source-item pair of the profile, that can be found out as $(0.2T, 0.8\ C_T)$.
3. Now the distance (Euclidean distance) between each indicator-specific source-item pair and the ideal 80/20 rule reflecting source-item pair can be found out using:

$$d(I-x)=\sqrt{(0.2T-x)^2+(0.8C_T-C_x)^2} \tag{1}$$

   Where x is the h-type indicator.

   For instance, if h is x, eqn. (2) becomes

$$d(I-h)=\sqrt{(0.2T-h)^2+(0.8C_T-C_h)^2} \tag{2}$$

4. Once all the distances for all the indicators are found for all the profiles in the data-set, then the average distance can be computed using

$$D(I-x)=\frac{\sum_i^N d_i(I-x)}{N} \tag{3}$$

   Where $d_i(I-x)$ represent the distance between an indicator pair for indicator x and the ideal indicator pair for profile i and N is the number of profiles.
5. The indicator pair with the least $D(I-x)$ is the closest (among the indicator pairs) to the ideal indicator pair. Let k be that indicator. Let n be the number of profiles for which the indicator pair for indicator k is the closest to the ideal indicator pair.

With this, indicator k can be found as the best among the indicators analyzed.

Objective 2: To investigate whether using 20 % of total publications, i.e., $0.2T$ as an indicator offers a better alternative, the following steps can be followed.

1. Find out $(y, C(y))$ indicator pair that corresponds to $0.2T$ as the indicator. Let $y=\max\ \{\lfloor 0.2T \rfloor,1\}$ represent the indicator and $C(y)$ be the citations received by top y publications. The max function is used keeping '1' as the other argument because, in some cases, $0.2T$ may be less than 1. In such cases, $C(y)=C(1)$, the top-most citations in a profile. The floor function is used to obtain y as an integer.
2. Now, similar to step 3 in the case of objective 1, find the distance between (I,C_I) and $(y, C(y))$ for all the profiles using

$$d(I-y)=\sqrt{(0.2T-y)^2+(0.8C_T-C_y)^2} \tag{4}$$

3. Once distances for all profiles are computed, the average distance can be computed using

$$D(I-y)=\frac{\sum_i^N d_i(I-y)}{N} \tag{5}$$

Where $d_i(I-y)$ represent the distance between the y indicator pair and the ideal indicator pair for a profile i.

4. If $D(I-y)<D(I-k), y$ is a better indicator than k.
5. Additionally, if the number of profiles for which indicator pair for indicator y is the closest to the ideal indicator pair will be greater than it is in the case of k. With the inclusion of indicator pair y for analysis, the number of profiles for which indicator pair of indicator k is the closest to the ideal indicator pair will change from n to n'. Let m, be the number of profiles for which the indicator pair for indicator y is the closest to the ideal indicator pair.

   Then, if $m > n'$,

   y, i.e., 0.2 T can better reflect the 80/20 law than indicator k (best among h-type indicators).

Objective 3*:* Though conditions stated in steps 4 and 5 (of procedure of objective 2) indicate that y is better indicator than k (and all other indicators analyzed), to test whether it has substantial capability to reflect the 80/20 rule can be determined in the following way.

1. Fix a rational threshold t_n.
2. If $\frac{m}{N}>t_n$, indicator y can substantially reflect the 80/20 law.

Though it is difficult to specify the value of t_n to be used, as a rule of thumb we recommend 0.5. The rationale is that if the indicator pair for y is found closest to the ideal indicator pair in more than 50% of the profiles it can be deemed to be capable of substantially reflecting the 80/20 law.

If any of the analysed indicators cannot reflect the 80/20 law then there is either a need for a new indicator or a new law (based on which the best base indicator can be determined).

4. Results and Discussion

Objective 1: As per steps 1 and 2, all the indicator pairs and the ideal indicator pair are computed. These are given in Table 1. After that, following step 3, $d(I-x)$ for all indicators are computed and are shown in Table 2 and visualized in Figure 1**.** A very close observation of Figure 1 (and also Table 2) will make it clear that all the indicators have shown the least deviation from the ideal indicator pair at least once.

This indicates the capability of each indicator to reflect the 80/20 rule (at least in some real-world profiles). However, to choose the best out of these, steps 4 and 5 have to be followed.

As per step 4, $D(I-x)$ for all the indicators are also computed and given in Table 3.

Table 1
Indicator-citation pairs of h-type indicators, the ideal indicator-pair and 'y' pair for 50 distinguished reviewers of 'Scientometrics' (profiles used by Lathabai [14])

Researcher	h-type indicator pairs												Ideal pair		y pair	
	T	C_T	$h(2)$	$C_{h(2)}$	h	C_h	f	C_f	t	C_t	g	C_g	I	C_I	y	C_y
Abramo	167	4422	10	1659	39	3372	51	3746	55	3839	63	4000	33.4	3537.6	33	3120
Adams	145	3964	9	1612	36	3154	48	3489	52	3564	60	3691	29	3171.2	29	2864
Aksnes	141	3164	10	2158	25	2839	33	2966	38	3008	55	3101	28.2	2531.2	28	2905
Archambault	74	3438	10	2473	24	3204	32	3319	39	3376	58	3435	14.8	2750.4	14	2831
Augillo	305	3686	9	1696	31	2809	43	3130	48	3234	58	3394	61	2948.8	61	3437
Bensman	38	937	6	479	18	854	22	899	25	918	30	934	7.6	749.6	7	524
Bordons	173	6108	11	3016	39	5126	54	5589	61	5736	77	5945	34.6	4886.4	34	4909
Boyack	171	7963	12	5133	38	7161	53	7601	61	7727	89	7923	34.2	6370.4	34	6697
Burrell	112	2100	8	866	25	1635	34	1809	37	1852	44	1938	22.4	1680	22	1557
Cabanac	138	1019	5	211	17	498	22	574	23	588	24	602	27.6	815.2	27	640
Chinchilla	239	3910	9	1629	30	2735	42	3039	47	3141	57	3300	47.8	3128	47	3141
Costas	186	5138	11	2797	33	4203	43	4574	47	4677	69	4867	37.2	4110.4	37	4327
D'Angelo	176	4563	10	1659	38	3332	50	3688	55	3801	62	3940	35.2	3650.4	35	3209
Daniel	260	8354	12	4627	41	6612	57	7188	66	7455	89	7934	52	6683.2	52	7023
Daraio	136	5447	11	3384	34	4827	46	5126	52	5205	73	5379	27.2	4357.6	27	4567
De Moya	520	10017	11	2753	53	6196	71	7007	76	7207	87	7605	104	8013.6	104	8101
Debackere	535	10846	13	5877	46	9364	63	9925	74	10131	102	10468	107	8676.8	107	10508
Ding	336	9922	12	4058	51	7699	69	8412	77	8632	95	9030	67.2	7937.6	67	8349
Gonzalez	36	812	5	630	12	755	15	786	18	798	28	812	7.2	649.6	7	687
Gorriaz	153	1597	7	590	22	1117	30	1269	32	1299	36	1347	30.6	1277.6	30	1269
Guan	291	9939	12	4407	49	7592	66	8242	74	8474	94	8924	58.2	7951.2	58	7976
Harzing	168	19241	19	10940	64	17709	87	18700	101	18991	138	19233	33.6	15392.8	33	14304
Haunschild	123	1705	7	585	24	1141	31	1282	33	1315	37	1377	24.6	1364	24	1141
Haunstein	103	5511	12	3957	29	5040	41	5302	49	5398	74	5506	20.6	4408.8	20	4691
Jamali	146	4997	10	2193	39	4095	52	4507	57	4613	69	4788	29.2	3997	29	3634
Kajikawa	199	4595	10	2159	35	3796	46	4058	53	4176	65	4317	39.8	3676	39	3904
Kousha	100	3586	10	1984	29	3119	41	3386	46	3445	59	3547	20	2868.8	20	2729
Lariveire	402	12620	14	5840	52	10198	75	11184	86	11492	109	11912	80.4	10096	80	11335

Contd...

Li	189	7615	10	4048	41	6211	56	6739	62	6904	85	7369	37.8	6092	37	6037
Lopez	27	622	6	490	11	579	15	612	17	618	24	622	5.4	497.6	5	439
Myer	176	8173	13	4646	41	7403	57	7872	66	7999	90	8131	35.2	6538.4	35	7092
Milojevic	85	2001	8	1089	21	1740	30	1886	35	1940	44	1982	17	1600.8	17	1614
Moutan	192	13574	10	11245	32	12459	46	12841	56	13029	116	13543	38.4	10589.2	38	12637
Park	163	7092	11	2285	48	5029	64	5685	69	5843	78	6085	32.6	5673.6	32	4119
Perc	383	25818	16	7312	94	19148	124	21500	133	22050	151	23036	76.6	20654.4	76	17300
Peters	144	2883	8	1722	23	2411	33	2593	38	2657	52	2783	28.8	2306.4	28	2512
Pouris	139	1969	7	645	24	1334	33	1518	36	1567	40	1625	27.8	1575.2	27	1403
Rafols	106	7882	12	5630	34	7372	47	7661	57	7801	88	7882	21.2	6305.6	21	6716
Robinson	123	1770	8	757	23	1343	32	1516	34	1543	40	1614	24.6	1416	24	1366
Schranhost	228	3169	8	1045	31	2191	41	2458	45	2535	51	2634	45.6	2535.2	45	2535
Schubert	126	2674	8	1152	26	1963	37	2211	41	2285	49	2410	25.2	2139.2	25	1935
Song	303	3586	8	1046	32	2116	41	2369	44	2440	50	2567	60.6	2868.8	60	2741
Sugimoto	182	7542	12	4037	43	6108	58	6626	65	6814	84	7170	36.4	6033.6	36	5790
Torgler	487	17892	17	8735	64	14602	86	15596	99	15990	129	16674	97.4	14313.6	97	15933
Van Den	295	4936	10	1947	39	3776	52	4196	57	4310	66	4462	59	3948.8	59	4348
Van Leuven	214	7984	12	3707	48	6816	63	7327	71	7495	87	7697	42.8	6387.2	42	6512
Vaughan	118	5081	10	2899	32	4465	45	4762	53	4895	71	5046	23.6	4064.8	23	4051
Waltman	55	7363	12	6201	26	7124	36	7296	45	7358	55	7363	11	5890.4	11	6055
Yan	97	2835	10	1483	29	2426	37	2591	41	2642	52	2744	19.4	2268	19	2036
Zhou	63	1735	7	1197	18	1561	24	1644	28	1675	41	1727	12.6	1388	12	1384

Table 2

Distances between the ideal indicator pair and *h*-type indicators and the *y*-pair for 50 distinguished reviewers of 'Scientometrics' (profiles used by Lathabai [14]. Bold values indicate least deviation from the ideal pair.

Researcher	*d(I-h(2))*	*d(I-h)*	*d(I-f)*	*d(I-t)*	*d(I-g)*	*d(I-y)*
Abramo	1878.75	**165.69**	209.14	302.17	463.35	417.6
Adams	1559.33	**18.57**	318.37	393.47	520.72	307.2
Aksnes	373.64	307.82	434.83	476.9	570.43	373.8
Archambault	277.44	453.69	568.86	626.07	685.96	**80.6**
Augillo	1253.88	**142.98**	182.09	285.5	445.2	488.2
Bensman	270.6	104.92	150.09	169.3	185.76	225.6
Bordons	1870.55	239.64	702.87	850.01	1059.45	**22.61**
Boyack	1237.6	790.61	1230.74	1356.86	1553.57	626.6
Burrell	814.13	**45.08**	129.52	172.62	258.9	123
Cabanac	604.62	317.38	241.27	227.25	213.23	**175.2**

Contd...

Chinchilla	1499.5	393.4	89.19	**13.02**	172.25	**13.02**
Costas	1313.66	**92.7**	463.7	566.82	757.27	216.6
D'Angelo	1991.56	318.41	**40.41**	151.9	290.84	441.4
Daniel	2056.59	**72.04**	504.82	771.93	1251.35	339.8
Daraio	973.73	469.45	768.63	847.76	1022.43	**209.4**
De Moya	5261.42	1818.32	1007.14	1474.03	408.95	**87.4**
Debackere	2801.38	**689.9**	1248.98	1454.57	1791.21	1831.2
Ding	3879.99	239.15	474.4	694.47	1092.75	411.4
Gonzalez	19.72	105.51	136.62	148.79	163.73	37.4
Gorriaz	688	160.83	**8.62**	21.45	69.61	**8.62**
Guan	3544.5	359.32	290.9	523.04	973.46	**24.8**
Harzing	4452.82	2316.4	3307.63	3598.83	3841.62	**1088.8**
Haunschild	779.2	223	82.25	49.71	**17.97**	223
Haunstein	451.88	631.26	893.43	989.61	1098.5	**282.2**
Jamali	1804.7	**97.89**	509.91	616.03	791.4	363.6
Kajikawa	1517.29	120.1	382.05	500.17	641.5	228
Kousha	884.86	250.36	517.63	576.79	679.32	**139.8**
Lariveire	4256.52	**105.88**	1088.01	1396.01	1816.23	1239
Li	2044.19	119.04	647.26	812.36	1277.87	**55.01**
Lopez	**7.62**	81.59	114.8	120.96	125.78	58.6
Myer	1892.53	864.62	1333.78	1460.92	1593.54	**553.6**
Milojevic	511.88	139.26	285.5	339.68	382.15	**13.2**
Moutan	**386.84**	1599.81	1981.81	2169.87	2684.92	1777.8
Park	3388.67	644.78	**33.41**	173.27	413.9	1554.6
Perc	13342.54	1506.5	**846.93**	1396.74	2382.76	3354.4
Peters	584.77	**104.76**	286.63	350.72	477.16	205.6
Pouris	930.43	241.23	57.44	**11.6**	51.27	172.2
Rafols	675.66	1066.48	1355.65	1495.83	1577.81	**410.4**
Robinson	659.21	73.02	100.27	127.35	198.6	**50**
Schranhost	1490.67	344.51	77.34	**0.63**	98.95	**0.63**
Schubert	987.35	176.2	**72.76**	146.65	271.84	204.2
Song	1823.56	753.34	500.18	429.12	301.99	**127.8**
Sugimoto	1996.75	**74.69**	592.79	780.92	1137.4	243.6
Torgler	5579.18	**290.33**	1282.45	1676.4	2360.61	1619.4
Van Den	2002.4	173.95	247.3	361.21	513.25	399.2
Van Leuven	2680.38	428.83	940.02	1108.16	1310.55	**124.8**
Vaughan	1165.88	400.29	697.53	830.72	982.34	**13.81**
Waltman	310.6	1233.69	1405.82	1467.99	1473.26	164.6
Yan	785.06	158.29	323.48	374.62	477.12	232
Zhou	191.08	173.08	256.25	287.41	340.19	**4.04**

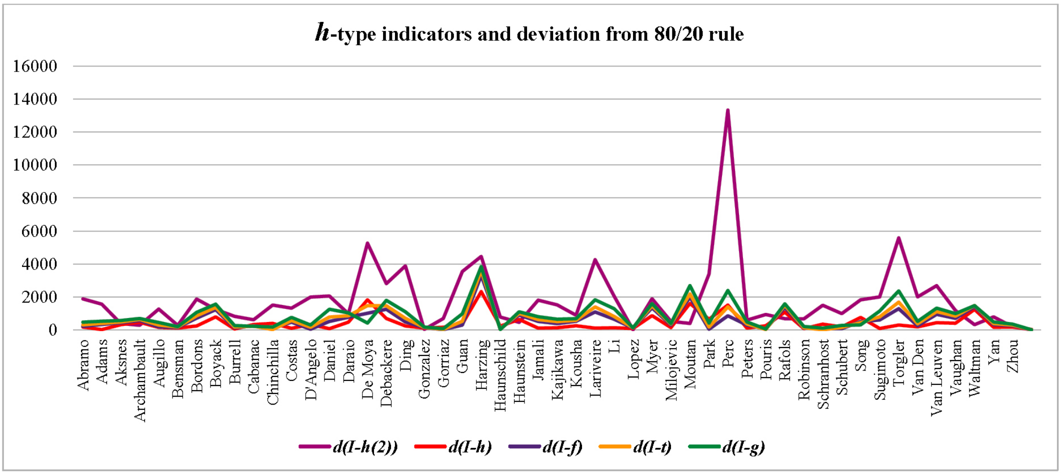

Figure 1

Deviations of *h*-type indicator pairs from the ideal pair

Table 3

***h*-type indicator pairs analysed and the number of profiles in which each indicator pair found closest to the ideal pair**

***h*-type indicator pair**	***D*(I – *x*)**	**# of profiles in which indicator pair best reflected 80/20 law**
$(h(2), C(h(2)))$	1835.102	7
$(h, C(h))$	**433.972**	**30**
$(f, C(f))$	588.43	6
$(t, C(t))$	703.564	3
$(g, C(g))$	865.404	4

As least $D(\mathrm{I} - x)$ is obtained for the *h*-index, it is the most suitable candidate among the indicators to reflect the 80/20 law. This can be reinforced by following step 5 in determining the number of profiles (from Table 2) in which each indicator pair are found closest to the ideal pair (i.e., number of profiles in which deviation is least for an indicator pair). The counts of profiles are given in Table 3. Here also, *h*-index outsmarts other indicators, with n=30. Thus, according to step 5, the *h*-index is the k in the group.

Thus, among the analyzed *h*-type indicators, the *h*-index is the best in reflecting the 80/20 law.

However, to test whether $0.2T$ can do any better, the procedure for objective 2 has to be followed, which is discussed next.

Objective 2: Following step 1, the indicator pair for y is computed and is given in the last two columns of Table 1. Following step 2, $d(\mathrm{I} - y)$ for all profiles are computed and shown in Table 2 (see last column). These are also shown in Figure 2 in comparison to the deviations

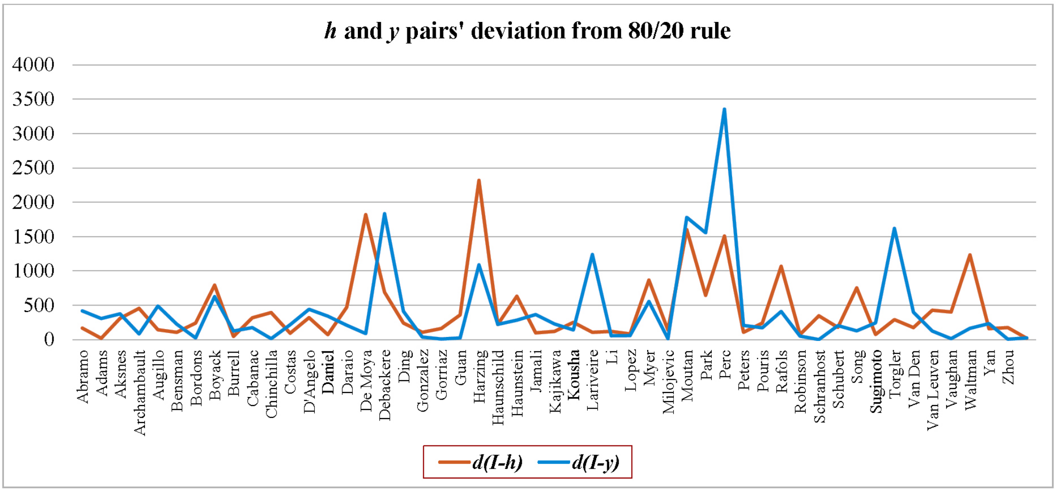

Figure 2

Deviations of *h* and *y* indicator pairs from the ideal pair

of the *h* pair. Deviations of all the *h*-type indicator pairs and *y*-pair from the ideal pair is given in Table 3.

By step 3, $D(\mathrm{I}-y) = 427.3068$

We have, $D(\mathrm{I}-h) = 433.972$ (from Table 3) and it does not change with the inclusion of *y* pair into analysis.

According to step 4, $D(\mathrm{I}-y) < D(\mathrm{I}-k) = D(\mathrm{I}-h)$, indicating 0.2$T$ indicator is slightly better than the *h*-index (but very close to it) in reflecting the 80/20 law.

Table 4

***h*-type and y indicator pairs and the number of profiles in which each indicator pair found closest to the ideal pair**

Indicator pair	# of profiles in which indicator pair best reflected 80/20 law
$(h(2), C(h(2)))$	3
$(h, C(h))$	18
$(f, C(f))$	5
$(t, C(t))$	3
$(g, C(g))$	1
$(y, C(y))$	**23**

To reinforce this, step 5 needs to be done. The inclusion of *y*-pair into the analysis changes the number of profiles in which each indicator pair best reflected the 80/20 law (i.e., last column of Table 3 changes). This change is indicated in Table 4.

From Tables 3 and 4, for the *h*-index, n=30 drops to n'=18. From Table 4, m=23.

Thus, $m > n'$. So, the *y*-indicator pair outperforms the indicator pair with respect to the *h*-index (though in a narrow margin). It can also be noted that in the case of least deviations, profiles where ties between *the y*-indicator and one or more of the *h*-type indicators exist. Thus, if $0.2T$ indicator is used as an indicator can marginally perform better than *h*-index and *h*-type indicators. However, to determine whether it can substantially reflect 80/20, procedure for objective 3 has to be conducted.

Objective 3: We fix the threshold, t_n at 0.5.

By step 2, $\frac{m}{N} = \frac{23}{50} = 0.46$

Thus, $\frac{m}{N} < t_n$.

This indicates that 0.2 T indicator (or any of the analyzed indicators) does not substantially reflect the 80/20 law. This is suggestive of the possibility of the existence of a better indicator to reflect the 80/20 law or the need for a new law itself. Thus, more rigorous explorations in these directions are needed either to develop a base indicator that can better reflect the 80/20 law or establish the suitability of a new law and efforts to find a base indicator with maximum compliance to the new law. However, few challenges are there for the discovery of new base indicator that can better reflect 80/20 law.

5. Challenges to develop new indicator

Though appeal of the new base indicator lies in (but not limited to) the above-mentioned possibilities, certain challenges are there to find that indicator or new law that may lead to that indicator. As all the *h*-type indicators are more or less dependent on the skewness of citation distributions reflected by the Lotkaian exponent (α), to determine the best indicator or best law that can be applied to a dominant majority of cases in real-world IPPs like paper-citation systems, the relationship between the indicator/law and α needs to be clearly established. Additionally, the new indicators should be located among the existing relationships (especially the order relationship) between *h*-type indicators like $h \le f \le t \le g$ [13], $h \le g \le \psi$ [14] (that together allows us to deduce that $h \le f \le t \le g \le \psi$) and $h \le p \le g \le \psi$ [31], where p is the *p*-index [32], [33]. However, some factors may hinder this. The major one is discussed below.

5.1 *Challenges in investigation of order relationship*

Unlike in the case of the above-mentioned inconsistencies, the order relationship between *h*-type indicators like $h \le f \le t \le g \le \psi$ and $h \le p \le g \le \psi$ given by theoretical frameworks are practically found to be very much valid. However, in the case of the *Ideal indicator*, from our minimalistic empirical exercise itself, it is found that different indicators can get closer to the 'ideal' behaviour (i.e., reflecting 80/20 law). The graphical visualization of the ideal indicator, *h* and *f* (*h* and *f* are chosen because after *y*-indicator, these exhibits of ideal behaviour in highest number of chosen profiles) is shown in Figure 3. This clearly indicates that ideal indicator can fall below *h* (that is why in some profiles *h(2)* exhibits ideal behaviour), go higher than *f* (that makes *t* and *g* to follow ideal behaviour in some of the profiles). Specially note that we have not included 0.2*T* or *y*-index in Figure 3 because it is almost same as that of the *I*-index and is difficult to distinguish. This definitely indicates that it is difficult for new base indicator (which is expected to exhibit ideal behaviour most of the times) to have a firm or permanent location in the citation curve. However, if the base indicator can only be developed from a different law, this may not be the case.

Once such an indicator is in place, further explorations that may lead to more sophisticated indicators that address problems like multiple authorship, field bias, etc. can be pursued using that base indicator.

Apart from this, having a better base indicator may contribute in the following ways to the field of scientometrics/informetrics and beyond apart from its use for productivity assessment of actors of science and technology and hence explorations in following directions can also be considered.

6. Possible applications of improved base indicator

6.1 *Scientometric and STI management applications*

h-index and its variants (base indicators) has inspired or influenced the development of indicators within the realm of scientometrics and beyond. For instance, Lathabai et al. [34] developed expertise indicators like *x* and *x(g)* (inspired by *h* and *g* respectively) that could reflect core competencies and potential core competencies of actors like institutions within

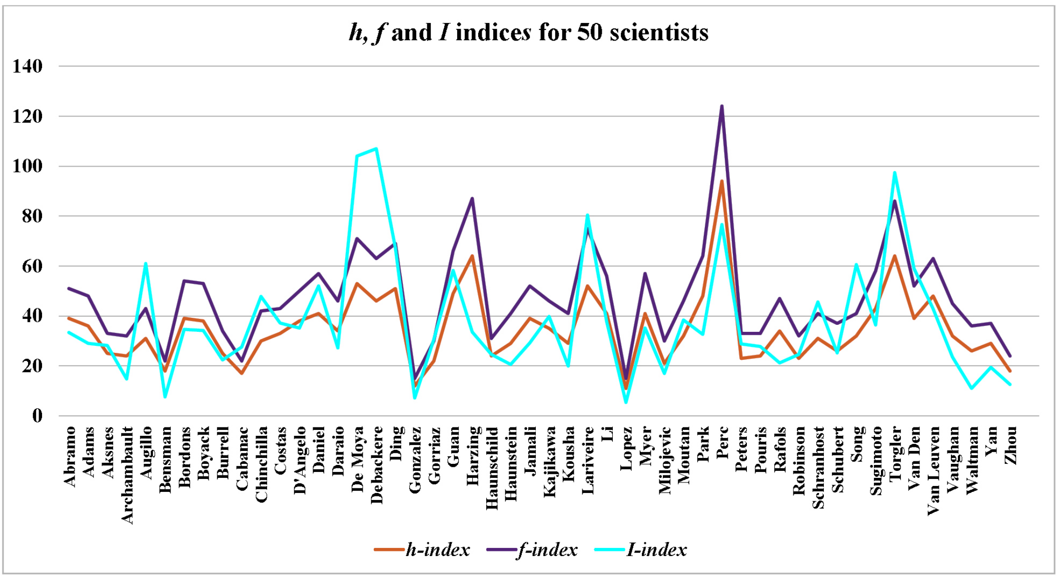

Figure 3

***h*, *f* and *I* indices for 50 scientists indicating the difficulty of determining the order relationship of new base indicator with other *h*-type indicators**

a discipline. An expertise-based collaboration recommendation system framework that can be useful for research portfolio management of institutions and national scholarly ecosystem management of countries with respect to a discipline was also contributed by [35], [36], [37], [38] developed expertise-diversity indicator viz. x_d -index (again computed in *h*-like fashion) that reflects the overall or broad-level core competences of institutions and thereby reflects the diversity of the research portfolio of institutions. This can be used to estimate the overall core competence and diversity of scholarly ecosystem of a country. These indicators and frameworks can be vital to Science, Technology and Innovation (STI) policymakers for the STI management of the country. If a better base indicator is there, suitable improvements to the *h*-index inspired fine-level expertise indicator x and broad level overall expertise-diversity indicator like x_d can be done accordingly to have better indicators and betterment of the recommendation system framework.

6.2 *Applications outside the realm of scientometrics*

Inspired by *h*-index, researchers have also developed some network indicators that can be used for a myriad of applications due to the appeal of networks in many real-world problems (that again include scientometric/informetric applications too). Lobby index or *l*-index, a node centrality measure was developed by Korn et al. [39], which can be computed in *h*-like fashion. The high level of congruence between definitions of *h*-index in case of citation profile of an actor (paper-citation system) and *l*-index of a node in a network (just like the case of *h*-index and expertise indicators discussed in section 6.1) aided the adoptability and applicability of *h*-index in the context of networks. *l*-index is found to work more efficiently than some other node centrality measure in linguistic networks [40] for ranking purposes. *wl*-index or weighted lobby index [41] is found to be better than *l*-index for analysis of stock market networks for development of stock recommendation and portfolio management system. For these kinds of applications, instead of *h*-index, better base indicator (once it is unveiled) can be used. Due to ubiquity of networks, such a base indicator can have potential applications in social, technological, biological and information (that include scientometric/informetric applications) systems that can be modelled as networks.

7. Conclusion

Owing to the skewness in source-item systems or IPPs like paper-citation systems, the 80/20 law might be at work. Though most of the prominent indicators like *h*-type indicators in paper-citation systems (applicable in many other source-item systems too) are also dependent on skewness, the connection between the 80/20 law and these indicators had not been explored properly. This is addressed in this work through the stated investigation points and we found that (i) all the *h*-type indicators have ability to reflect 80/20 law in at least some of the real-world profiles, but *h*-index is best among them, (ii) $0.2T$ is slightly better than *h*-index in reflecting 80/20 law, but (iii) $0.2T$ cannot be considered as the most suitable indicator that can substantially reflect 80/20 law in a majority of real-world profiles (source-item systems). All these points towards the need for and possibility of the existence of a better base indicator to properly reflect the 80/20 law or perhaps the existence of a better law upon which a better indicator can be built up. However, the current study needs to be validated using much larger real-world dataset.

Investigations to discover such a base indicator or law can be attempted by analytical approach, empirical explorations, or both. Generalized law and methods to determine inequality for comparison of two cases as given by Rousseau et al. [3], [30] in case of author-article systems can be of help in explorations to devise indicator that ensures law-indicator compatibility. Major challenges that might be encountered in analytical as well as empirical studies in this regard are also highlighted in this work. More challenges might be out there. Tracking and addressing those can also be considered. Sophisticated methods including those under the umbrella of machine learning can be considered for this purpose, especially for the case of investigation of the existence of a new law. We speculate that by employing machine-learning approaches on much larger datasets, the pattern or nature of the relationship between sources and items could be revealed (training) and validated too on substantially large datasets (testing). However, the feasibility of this can be confirmed only with further explorations. If the new indicator or law is determined using any other approaches, apart from validation of the current study, there is a need for testing of new indicator thus identified using much larger real-world datasets. These are some of the possible further endeavours that may entice scientometricians. Once new base indicator (which is close to ideal indicator) is in place, several indicators that are built using the principle of *h*-index for applications within the realm of scientometrics and beyond (that we discussed in section 6 and more) can be explored.

Declarations

Conflict of interests

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

Funding

The authors declare that the manuscript is not funded or supported by any of the funding organizations or schemes.

References

[1] V. Pareto, "Coursd'économie politique: Professé à l'Universi", *té de Lausanne*, vol. 1 (1926).

[2] L. Egghe, "The power of power laws and an interpretation of Lotkaian informetric systems as self-similar fractals," *Journal of the American Society for Information Science and Technology*, vol. 56, no. 7, pp. 669–675 (2005), doi: https://doi.org/10.1002/asi.20158.

[3] "The Informetric Laws". In R. Rousseau, L. Egghe, and R. Guns (Eds.), Becoming metric-wise, pp. 293-313 (2018).

[4] Q.L. Burrell. "The 80/20 rule: Library lore or statistical law?". *Journal of Documentation*, vol. *41*, no. 1, pp. 24-39 (1985).

[5] L. Egghe. "On the 80/20 rule". *Scientometrics*, vol. 10, no. 1, pp. 55-68 (1986).

[6] L. Egghe, "Pratt's measure for some bibliometric distributions and its relation with the 80/20 rule". *Journal of the American Society for information Science*, vol. 38, no. 4, pp. 288-297 (1987).

[7] L. Egghe. "An exact calculation of Price's law for the law of Lotka". *Scientometrics*, vol. 11, no. 1-2, pp. 81-97 (1987).

[8] J. E. Hirsch, "An index to quantify an individual's scientific research output," *Proceedings of the National Academy of Sciences*, vol. 102, no. 46, pp. 16569–16572 (Nov. 2005), doi: https://doi.org/10.1073/pnas.0507655102.

[9] L. Egghe and R. Rousseau, "An informetric model for the Hirsch-index," *Scientometrics*, vol. 69, no. 1, pp. 121–129 (Oct. 2006), doi: https://doi.org/10.1007/s11192-006-0143-8.

[10] M. Kosmulski, "A new Hirsch-type index saves time and works equally well as the original h-index". *ISSI newsletter*, vol. 2, no. 3, pp. 4-6 (2006).

[11] L. Egghe, "An improvement of the h-index: the g-index," vol. 2, no. 1, pp. 8–9 (Jan. 2006).

[12] L. Egghe, "Theory and practise of the g-index," *Scientometrics*, vol. 69, no. 1, pp. 131–152 (Oct. 2006), doi: https://doi.org/10.1007/s11192-006-0144-7.

[13] R. S. J. Tol, "The h-index and its alternatives: An application to the 100 most prolific economists," *Scientometrics*, vol. 80, no. 2, pp. 317–324 (Apr. 2009), doi: https://doi.org/10.1007/s11192-008-2079-7.

[14] H. H. Lathabai, "ψ-index: A new overall productivity index for actors of science and technology," *Journal of Informetrics*, vol. 14, no. 4, p. 101096 (Nov. 2020), doi: https://doi.org/10.1016/j.joi.2020.101096.

[15] F. L. da Silva, L. C. Brandão, and J. C. C. B. Soares de Mello, "Stability Discussions on some h-type Indexes," *Journal of Scientometric Research*, vol. 10, no. 1, pp. 9–17 (May 2021), doi: https://doi.org/10.5530/jscires.10.1.2.

[16] L. Bornmann, R. Mutz, and H.-D. Daniel, "The h index research output measurement: Two approaches to enhance its accuracy," *Journal of Informetrics*, vol. 4, no. 3, pp. 407–414 (Jul. 2010), doi: https://doi.org/10.1016/j.joi.2010.03.005.

[17] W. Glänzel, "On the opportunities and limitations of the H-index", *Science focus*, vol. 1, no. 1, pp. 10-11 (2006).

[18] W. Glänzel and H. F. Moed, "Opinion paper: thoughts and facts on bibliometric indicators," *Scientometrics*, vol. 96, no. 1, pp. 381–394 (Nov. 2012), doi: https://doi.org/10.1007/s11192-012-0898-z.

[19] L. Leydesdorff, L. Bornmann, and T. Opthof, "hα: the scientist as chimpanzee or bonobo," *Scientometrics*, vol. 118, no. 3, pp. 1163–1166 (Jan. 2019), doi: https://doi.org/10.1007/s11192-019-03004-3.

[20] X. Hu, R. Rousseau, and J. Chen, "In those fields where multiple authorship is the rule, the h-index should be supplemented by role-based h-indices," *Journal of information science*, vol. 36, no. 1, pp. 73–85 (Nov. 2009), doi: https://doi.org/10.1177/0165551509348133.

[21] M. Schreiber, "A modification of the h-index: The hm-index accounts for multi-authored manuscripts," *Journal of Informetrics*, vol. 2, no. 3, pp. 211–216 (Jul. 2008), doi: https://doi.org/10.1016/j.joi.2008.05.001.

[22] S. Alonso, F. J. Cabrerizo, E. Herrera-Viedma, and F. Herrera, "h-Index: A review focused in its variants, computation and standardization for different scientific fields," *Journal of Informetrics*, vol. 3, no. 4, pp. 273–289 (Oct. 2009), doi: https://doi.org/10.1016/j.joi.2009.04.001.

[23] N. J. Van Eck, and L. Waltman, "Generalizing the h-and g-indices", *Journal of Informetrics,* vol. 2, no. 4, pp. 263-271 (2008).

[24] L. Bornmann and L. Leydesdorff, "Count highly-cited papers instead of papers with h citations: use normalized citation counts and compare 'like with like'!," *Scientometrics*, vol. 115, no. 2, pp. 1119–1123 (Mar. 2018), doi: https://doi.org/10.1007/s11192-018-2682-1.

[25] L. Waltman, and N. J. Van Eck, "The inconsistency of the h-index", *Journal of the American Society for Information Science and Technology,* vol. 63, no. 2, pp. 406-415 (2012).

[26] J. E. Hirsch, "hα: An index to quantify an individual's scientific leadership," *Scientometrics*, vol. 118, no. 2, pp. 673–686 (Jan. 2019), doi: https://doi.org/10.1007/s11192-018-2994-1.

[27] D. J. Urquhart, "Urquharts Law", *Journal of Documentation*, vol. *33*, no. 2, pp. 149-149 (1977).

[28] E. Garfield, "The mystery of the transposed journal lists—wherein Bradford's Law of Scattering is generalized according to Garfield's Law of Concentration". *Essays of an information scientist*, vol. 1, pp. 222-223 (Aug, 1971).

[29] S. J. Bensman, "Urquhart's and Garfield's Laws: The British controversy over their validity," *Journal of the American Society for Information Science and Technology*, vol. 52, no. 9, pp. 714–724 (2001), doi: https://doi.org/10.1002/asi.1122.

[30] "Statistics". In R. Rousseau, L. Egghe, and R. Guns (Eds.), Becoming metric-wise, pp. 67-97 (2018).

[31] S. X. Wei, T. Tong, R. Rousseau, W. Wang, and Y. Y. Fred, "Relations among the h-, g-, ψ-, and p-index and offset-ability". *Journal of Informetrics*, vol. 16, no. 4, pp. 101340 (2022).

[32] G. Prathap, "Is there a place for a mock h-index?," *Scientometrics*, vol. 84, no. 1, pp. 153–165 (Jun. 2009), doi: https://doi.org/10.1007/s11192-009-0066-2.

[33] G. Prathap, "The 100 most prolific economists using the p-index," *Scientometrics*, vol. 84, no. 1, pp. 167–172 (Jun. 2009), doi: https://doi.org/10.1007/s11192-009-0068-0.

[34] H. H. Lathabai, A. Nandy, and V. K. Singh, "x-index: Identifying core competency and thematic research strengths of institutions using an NLP and network based ranking framework," *Scientometrics*, vol. 126, no. 12, pp. 9557–9583 (Nov. 2021), doi: https://doi.org/10.1007/s11192-021-04188-3.

[35] H. H. Lathabai, H. H., A. Nandy, and V. K. Singh, "Expertise-based institutional collaboration recommendation in different thematic areas", *In proceedings of the 11th International Workshop on Bibliometric-enhanced Information Retrieval co-located with 43rd European Conference on Information Retrieval* (ECIR 2021), pp. 45-52 (2021).

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

[37] A. Nandy, H. H. Lathabai, and V. K. Singh, "xd-index and its variants: a set of overall scholarly expertise diversity indices for the research portfolio management of institutions", *Scientometrics*, vol. 129, pp. 5937–5962 (2024).

[38] A. Nandy, H. H. Lathabai, and V. K. Singh, "xd-index: an overall scholarly expertise index for the research portfolio management of institutions". *In proceedings of 19th International Conference of the International Society for Scientometrics and Informetrics Bloomington, IN, USA*, pp. 473-488 (2023).

[39] A. Korn, A. Schubert, and A. Telcs, "Lobby index in networks," *Physica A: Statistical Mechanics and its Applications*, vol. 388, no. 11, pp. 2221–2226 (Jun. 2009), doi: https://doi.org/10.1016/j.physa.2009.02.013.

[40] M. G. Campiteli, A. J. Holanda, L. D. H. Soares, P. R. C. Soles, and O. Kinouchi, "Lobby index as a network centrality measure," *Physica A: Statistical Mechanics and its Applications*, vol. 392, no. 21, pp. 5511–5515 (Nov. 2013), doi: https://doi.org/10.1016/j.physa.2013.06.065.

[41] S. George, H. H. Lathabai, T. Prabhakaran, and M. Changat, "Towards stock recommendation and portfolio management systems using network analysis," in *Transactions on computational science and computational intelligence*, pp. 203–215 (2021).