A career in numbers

A citation network analysis of the work of RP Millar and his contribution to GnRH research

Citation for published version:
Leng, R & Leng, G 2024, 'A career in numbers: A citation network analysis of the work of RP Millar and his contribution to GnRH research', *Journal of Neuroendocrinology*, pp. 1-12. https://doi.org/10.1111/jne.13430

Digital Object Identifier (DOI):
10.1111/jne.13430

Published In:
Journal of Neuroendocrinology

Received: 17 April 2024 | Revised: 11 June 2024 | Accepted: 25 June 2024

DOI: 10.1111/jne.13430

ORIGINAL ARTICLE

Journal of Neuroendocrinology WILEY

A career in numbers: A citation network analysis of the work of RP Millar and his contribution to GnRH research

Rhodri I. Leng[1] | Gareth Leng[2]

[1]School of Social Science and Political Science, University of Edinburgh, Edinburgh, UK

[2]Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh, UK

Correspondence
Gareth Leng, Centre for Discovery Brain Sciences, University of Edinburgh, Edinburgh, UK.
Email: gareth.leng@ed.ac.uk

Abstract

Here, we reflect on the long career in neuroendocrinology of a single, highly productive scientist ('Bob' Millar), by analysing his oeuvre of published papers through the lens of citation metrics. We use citation network analysis in a novel manner to identify the specific topics to which his papers have made a particular contribution, allowing us to compare the citations of his papers with those of contemporary papers on the same topic, rather than on the same broad field as generally used to normalise citations. It appears that citation rates are highest for topics on which Bob has published a relatively large number of papers that have become core to a tightly-knit community of authors that cite each other. This analysis shows that an author's impact depends on the existence of a receptive community that is alert to the potential utility of papers from that author, and which uses, amplifies, extends and qualifies the contents of their papers—activities that entail reciprocal citation between authors. The obvious conclusion is that a scientist's impact depends on the use that his or her contemporaries make of his or her contributions, rather than on the contributions in themselves.

KEYWORDS

citation network, GNIH, Kisspeptin, reproduction, science of science

1 | INTRODUCTION

Scientists generally know well the misuses to which citation metrics can be put. Nevertheless, they generally pay close attention to their own citation metrics, being keen to know that their own work, generated with such pain and angst, has been noticed. Here, we use the occasion of a Festschrift for Robert P ('Bob') Millar to examine the citations to his oeuvre—and the citations between his citing papers. Our purpose is neither to praise Bob nor to bury him in numbers, but to contextualise publication and citation metrics by a study that tries to evaluate the impact of a single individual through metrics that are commonly used to assess scientists.

Bob's research career has focussed on a molecular understanding of the regulation of GnRH actions; it has spanned more than 50 years, and has gained him a surfeit of scientific honours. We ask a simple question, what if anything, do citation metrics tell us about Bob's impact on his field? Citation and publication metrics are omnipresent and, despite their well-catalogued limitations,[1,2] many managers of science still use them to evaluate scientists and their institutions, and some scientists apparently use them to decide what papers are worth reading (and citing).

But how exactly do authors decide what papers to cite? There is a common assumption that how often researchers are cited reflects the *quality* of their work—that their citations reflect 'excellence and influence', in some senses of those terms. But Bruno Latour proposed a different answer to this question after spending a year in the laboratory of Roger Guillemin at the Salk Institute. There, he had assumed the role of an anthropologist living amongst the strange tribe of neuroendocrinologists. In *faux naivete*, he noted that while the scientists thought that their purpose was to discover facts, he 'doggedly argued that they were writers and readers in the business of being convinced and convincing others.' Bruno's account, co-written with Steve Woolgar, was published as '*Laboratory Life: The Construction of Scientific Facts*'.[3] It has been cited more than 25,000 times, and its evidently influential case impelled the conclusion that, when scientists write a paper, they select references not because of their quality, but because in some way they advance a claim that is asserted in that paper[4] (as also noted by Gilbert[5]). The strength of support for that claim might be greater when the reference is in a reputable journal, or from a well-known scientist, or when it is a recent contribution, or indeed if it contains strong evidence that supports the claim. On the other hand, papers that are relevant but which offer evidence that confuses or contradicts the claim are often neglected. And less well-known scientists are often neglected in favour of the well-known; low ranking journals in favour of the high ranking; and replication studies, however large and rigorous, in favour of the better-known early studies, however small or flawed they may be.[6–8]

At the time when *Laboratory Life* was published, there was pressure to measure the productivity of scientists, and perhaps it was natural to do so by counting the papers that they published. Soon, however, concerns were raised that this encouraged 'salami-slicing'—carving the outcomes of large studies into multiple elements that satisfied the condition of being 'the smallest publishable unit'.[9] Accordingly, in UK Research institutes, the metric was adjusted to count the number of *pages* published. However, scientists responded by publishing jointly, increasing the apparent output of everyone concerned, but without affecting the total output,[10] and perhaps as an unintended consequence, sole-author papers became an endangered species. The next step was to divide the number of pages in each paper by the number of authors, but since authorship was now irreversibly compromised by a culture of reciprocation, it was becoming clear that this metric had been destroyed by those on whom it had been inflicted. It began to look worryingly as though the only way to tell how good a paper was would be to read it.

However, bibliometric techniques seemed to offer respite from that refuge of last resort. If, as postulated by Merton,[11] citations acknowledged the debt that each paper owed to its predecessors, then counting citations would measure a paper's worth. But problems were apparent from the start of its application to assessment of science.[12] Methodological papers were sometimes cited for what seemed slight contributions: one paper[13] has been cited more than 40,000 times for proposing a questionnaire that can be used to determine whether someone is right-handed or left-handed (see[8]). Citation counts varied by research field,[14] in part due to different referencing conventions; for example, in the mathematical sciences, papers typically contained few references, and citations were correspondingly low. The variability also reflected the age of the field: in new, rapidly growing fields, citation counts were typically high simply because there were fewer relevant papers that might be cited.

But, nevertheless, citations hold interest for the 'science of science'.[15] Citation networks can capture aspects of the flow of ideas and the formal use of previous works—and the rise, fall and reconstruction of communities in a field. Thus, understanding how citations are used, by whom, and to what end, can help to understand the social structure of a field and the evolution of its theories and dogmas. Applying it to neuroendocrinology, we have used citation network analysis to inform a study of the evolution of the oxytocin field,[16] to analyse a controversy about the role of vasopressin in memory, and to dissect the case for oxytocin as a 'social hormone'.[17]

We have already made it clear that we think that citation counts have no place in the assessment of scientific quality[8]; anyone who wants to comment on the quality of Bob's work should read his papers and use their brains.

2 | METHODS

All citation data described in this paper are derived the Web of Science Core Collection (WoS). From this, we identified 477 documents authored by Bob between 1970 and 2023. Accompanying this document is their citation history: they have been cited 21,677 times by 11,138 documents indexed in the WoS. On average, they have 45.4 citations, with a h-index of 77. Removing self-citations leaves 19,256 citations by 10,719 documents.

We constructed a citation network dataset from these documents and every WoS indexed document that cited them. In this dataset, 11,196 nodes represent documents and 115,834 edges between nodes represent citation links from a citing to a cited document. We removed 67 papers for which metadata was incomplete or corrupted, and focussed on the largest interconnected component, removing nodes with no connections and clusters detached from the main network. Finally, as the WoS Core Collection includes only a small proportion of all Abstracts presented at Scientific Meetings, and, as they are rarely cited, contain few references, and their content is often duplicated in full papers, we excluded them, along with items such as letters, corrections and editorials. To do so, we removed papers with <10 references, reducing the dataset to 10,901 nodes and 113,742 edges, of which 389 are papers authored by Bob. Searches in WoS excluded Meetings Abstracts, other minor items and papers not published in English. Topic and title searches on WoS used the specified term and related terms as identified by WoS.

For each document, we established the *indegree* (how often it was cited by other documents in the network), the *outdegree* (how many of its references were to other documents in the network), and the *degree*—the sum of indegree and outdegree, a measure of its connectedness with other documents in the network. We also parsed in metadata from WoS, including the total number of citations and references per document. We clustered the network by modularity maximisation (a density-based clustering technique) using the Leiden algorithm.[18] Modularity maximisation is used to measure how well a

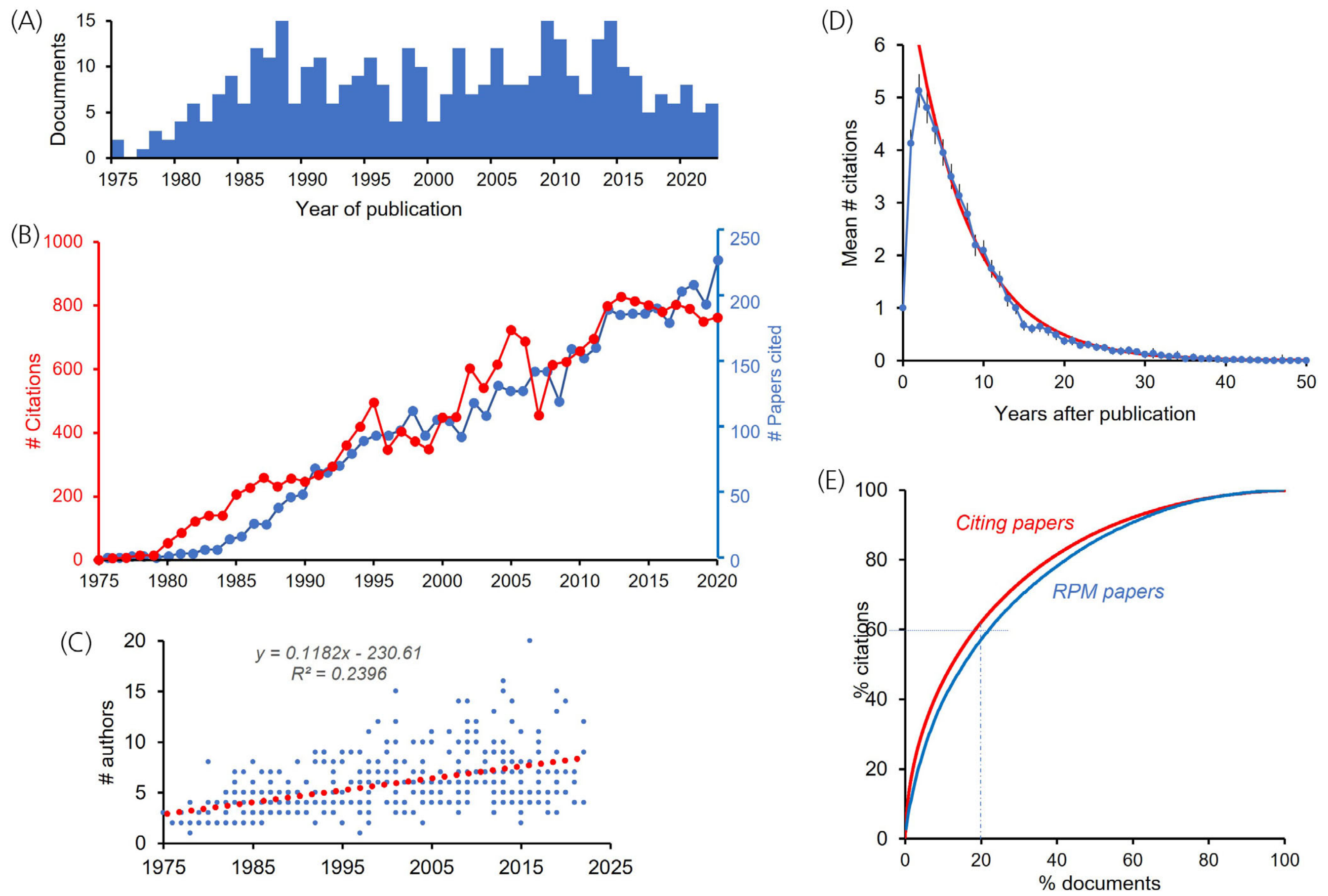

FIGURE 1 (A) Papers published by RP ('Bob') Millar by year. (B) The number of citations per year to Bob's papers (red) and the number of his papers that were cited each year (blue). (C) The number of authors in each of Bob's papers analysed: The red line is the linear trendline indicating an approximate doubling in the mean number of authors per paper over the time period. (D) Citations per year since publication for Bob's papers; blue lines are means± SEM. The red line is a negative exponential fitted to the data from 2 years after the date of publication. (E) Normalised distribution of citations for Bob's papers (blue) and for citing papers (red). Both distributions reflect the fact that a small proportion of papers attract very many citations; about 15% of Bob's papers receive 50% of all citations to them, and about 12% of the citing papers receive 50% of all citations to them.

network can be divided into 'modules' (clusters), and modularity (Q) is calculated by comparing the density of edges within communities to the expected density of edges for a random network with the same degree distribution. The goal of modularity maximisation is to partition a network into communities of densely connected nodes, with the nodes belonging to different communities that are comparatively sparsely connected. This process led to the identification of 25 clusters at $Q = 0.667$ (Resolution $= 1$; Number of iteration and restarts: 100). We parsed cluster membership data into the dataset, and annotated the clusters by reading a sample of the titles and abstracts of papers within each cluster. We then visualised this network (and subnetworks) with nodes positioned via a force-directed algorithm designed for the visualisation of cluster structure in large networks; this algorithm pulls together nodes that share edges, and pushes nodes apart that do not (see[19] for more details).

To identify each cluster, we looked at the citing papers with the highest degree—that is, those with most links (references and/or citations) to other papers in the network, having established (see below) that most of these links are to and from other papers in the same cluster. We looked at terms that appeared in the title of papers, as these are generally particularly salient for the author, whereas keywords often allude to a broader context. We sought to identify terms that distinguished between clusters, but two clusters (C3 and C8) both had 'GnRH' as the most common title term. Looking at these papers closely it was evident that C3 was mainly concerned with the variation of isoforms of GnRH amongst different species, while C8 was concerned with regulation of its expression.

3 | RESULTS

3.1 | Citation metrics

The 389 retained documents authored by Bob (with 925 different co-authors from 39 countries) comprise a segue from a sole author 1972 paper on 'Degradation of spermatozoa in [...] rock hyrax'[20] to a

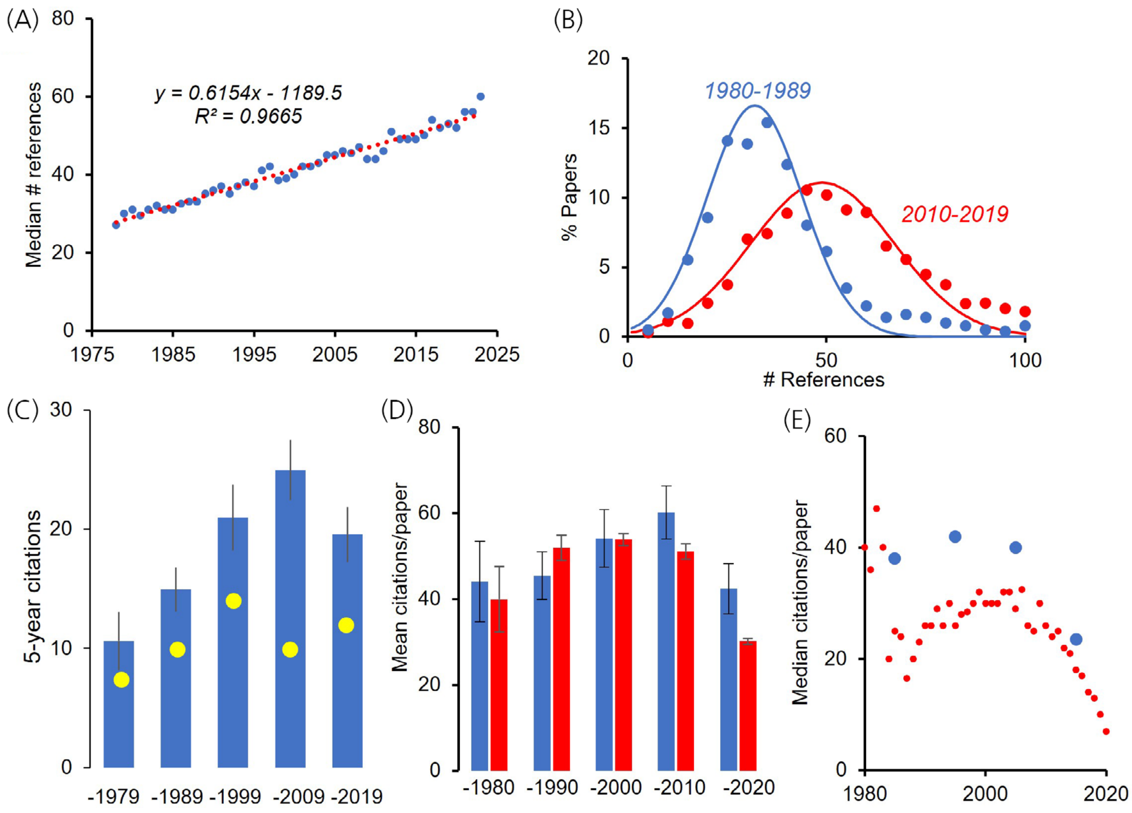

FIGURE 2 (A) Mean number of references in the citing papers by year of publication. The red line shows a linear fit to these data with the equation and R^2 value as shown. The mean size of reference lists has grown progressively over time, accounting for a corresponding increase in the mean citation rate of papers. (B) Distributions of reference list sizes for citing papers published in 1980–1989 (blue) and 2010–2019 (red); the data are normalised to the number of papers with <100 references. (1103 citing papers were published between 1980 and 1989, of these, 993 (85%) had <100 references, and these were mainly research papers. 3781 citing papers were published between 2010 and 2019, of these, 2966 (78%) had <100 references.) In both decades, the size of reference lists follows an approximately normal distribution for papers with <100 references. (C) Mean ± SEM number of citations in the 5 years after publication for Bob's papers published in each of the five decades up to January 2019. Medians are shown in yellow. The large discrepancy between mean and median values reflects a highly skewed distribution of citations. (D) Mean ± SEM citations to Bob's papers (blue) and citing papers (red) by decade of publication. (E) Red circles: Median citations per citing paper by year of publication; Blue circles: Median citations for Bob's papers by decade of publication. Note that while the mean citations to Bob's papers are similar to the mean citations to citing papers, the median citations to Bob's papers are much higher than the median citations to citing papers.

five-author paper in 2023 on 'Role of Neurokinin B in […] freshwater catfish'.[21] The rate of production of the documents was relatively stable between 1980 and 2022 (Figure 1A).

The rate of citations to Bob's papers increased linearly over time; paralleled by an increase in how many were cited each year; in each of the years 2010–2020, about 180 of his papers were cited at least once (Figure 1B). The average number of co-authors increased by about 2-fold (Figure 1C), but we found no apparent relationship between number of authors and citations.

Citations to Bob's papers peaked on average in the second year after publication, and thereafter the rate of citation declined with a half-life of about 5 years (Figure 1D).[22,23] The distribution of citations to his papers and that of citations to the cited papers both followed a long-tailed distribution, reflecting an unequal distribution of citations (Figure 1E): about 20% of Bob's papers receive about 60% of all citations to them, and the distribution is slightly more unequal for the citing papers. These distributions are similar to those for papers in the oxytocin field.[16]

Looking at the papers that cited Bob, the size of their reference lists increased from a mean (SEM) of 50 ± 2 in 1044 papers published before 1990 to 91 ± 2 in 1616 papers published since 2020, while the median increased from ~30 at the end of the 1970s to ~60 by 2020 (Figure 2A). The difference between mean and median reflects a skew in the distribution of reference list sizes introduced by an initially relatively small, but progressively increasing, proportion of reviews, which generally have many more references than research papers. The reference list sizes of papers with <100 references (i.e., mainly research papers) were distributed approximately normally both for papers published in 1980–1989 and for those published in 2010–2019 (Figure 2B). In the oxytocin literature, reference lists of research papers increased from a mean of 27 papers in 1980–1989 to 61 in 2010–2019, close to the median levels observed here.[16]

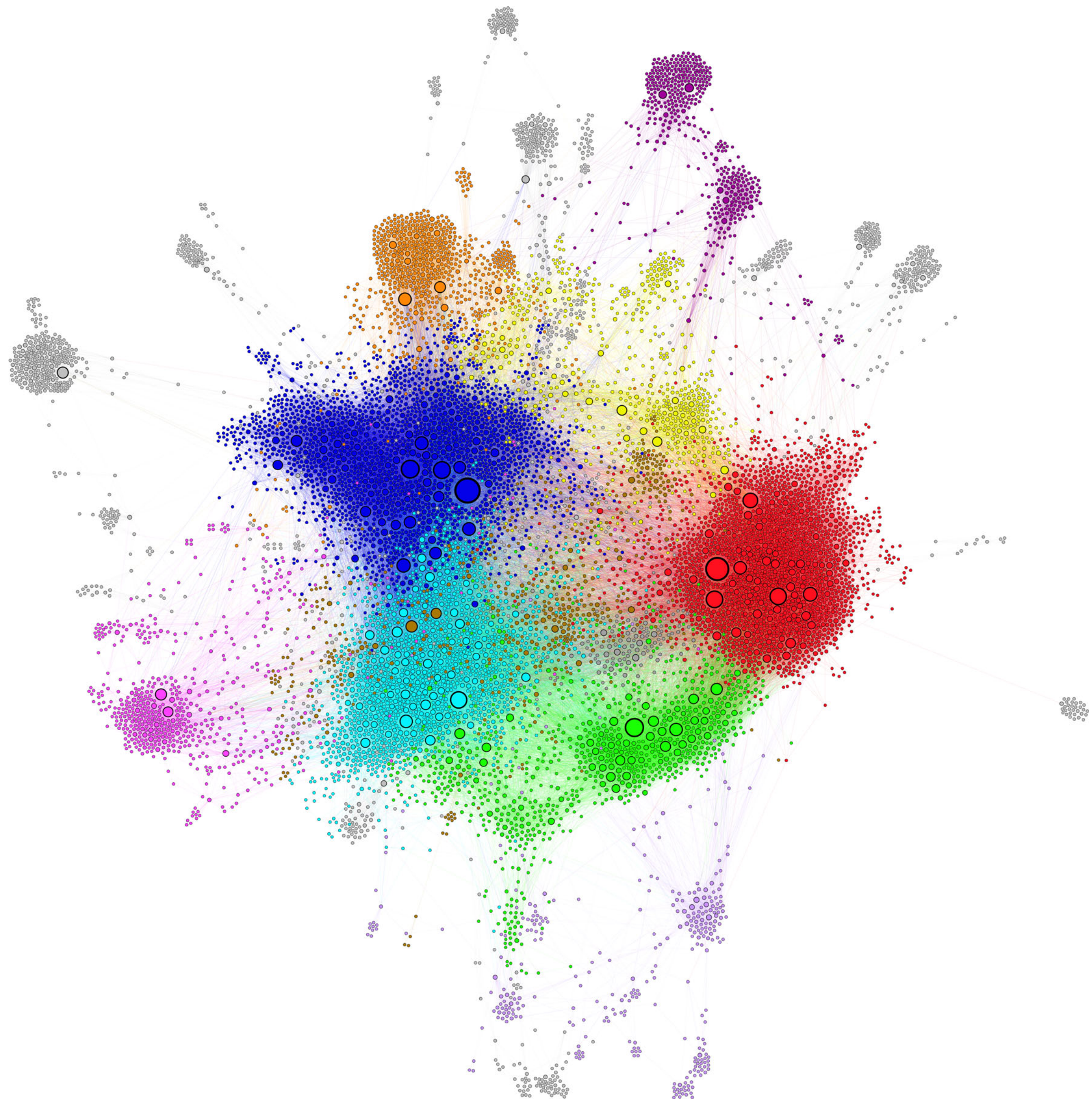

FIGURE 3 Citation network composed of 389 of Bob's papers and 10,512 papers that cited these (n = 10,901; m = 113,742). The node size is relative to indegree, and the node position was determined by the ForceAtlas2 algorithm. The colours are determined by cluster membership, as detailed in Table 1.

We then looked at the number of citations received by Bob's papers in the 5 years after publication by decade of publication (Figure 2C). The mean increased progressively in the three decades up to 2009, with a slight fall back in 2010–2019. The overall trend is a doubling in citations per paper over this period, as expected from the doubling in the size of reference lists in the same period. The median is consistently lower than the mean (as also observed for papers in the oxytocin field[16]), reflecting the skewed distribution of citations. However, the median shows little change with time, indicating that the increased mean reflects increasingly unequal rates of citation—as also observed in the oxytocin field.[16] Comparing Bob's papers with those that cited them (Figure 2E), there is little difference in the mean citations in any decade, but the median citations to Bob's papers is consistently higher (Figure 2F).

A casual interpretation of citation metrics is that they reflect the 'impact' that a paper has had. We could stop here, and just assume that the citation numbers alone tell us enough—but no thinking person should find this reasonable. To understand more, we have to understand what papers cited Bob and in what areas. Here, we use citation network analysis to identify clusters of interrelated papers in the network consisting of Bob's papers and their citing literature. This leverages the convention of referencing whereby scientists have cultivated the practical habit of referencing papers that are relevant to their papers. This creates a self-organising structure that pulls together papers into clusters that reflect specific topics, clusters interwoven by the references they send out and the citations they receive from other, similar papers.[24]

TABLE 1 Summary of the 25 clusters shown in Figure 3.

Cluster	Nodes	Bob's papers	Colour	Common keywords
1	2062	103	Blue	GnRH receptor; GnRH; GnRH-II
2	1960	57	Red	Kisspeptin; KISS1; neurokinin B; GPR54; KNDY
3	1327	62	Light Blue	GnRH; GnRH receptor; GnRH-II; olfactory placode
4	836	23	Green	GnIH; RFRP-3; GPR147; LPXRFA
5	729	12	Orange	Receptor activation; class B GPCR; chimeric receptor
6	609	18	Yellow	Kallmann syndrome; hypogonadotropic hypogonadism
7	588	15	Pink	Somatostatin, GH releasing factor; dynorphin; GH; LH
8	534	27	Brown	Neuropeptide K; GnRH-Associated Peptide; GnRH gene expression
9	485	10	Purple	EG-VEGF; prokineticins, BV8
10	321	14	Lilac	Mole-rat; reproductive suppression
11	245	1	Grey	Overtraining syndrome; underperformance; gymnastics
12	183	5	Grey	Ovulation induction; polycystic ovarian syndrome; hyperandrogenism
13	156	10	Grey	GnRH immunisation; immunocastration; immunocontraception
14	135	2	Grey	UTP; ATP; nucleotide receptor; purinoceptor
15	129	3	Grey	Endocrine-disrupting chemicals; polychlorinated biphenyls
16	126	13	Grey	Exocytosis; PLC; phosphatidylinositol; diacylglycerol
17	112	3	Grey	KISS2; GNRH1; GPR54; kisspeptin-10; chub; mackerel
18	101	3	Grey	TRH receptor; fluid secretion; Malpighian tubule; CRH; TSH
19	79	2	Grey	Medial habenula; Ca^{2+} signalling; mast cells; GnRH
20	66	1	Grey	Ion mobility-mass spectrometry; peptides; molecular dynamics
21	59	1	Grey	Mutation; DNA methylation; P53
22	31	2	Grey	African lion; bovine tuberculosis
23	13	1	Grey	Vasoinhibin; 16 K prolactin; angiogenesis
24	10	0	Grey	Porphyria
25	5	1	Grey	K562; HLA Class II

TABLE 2 Citation metrics of clusters C1–C10. The lead paper of each cluster (second column) was the citing paper with the largest degree.

		Bob's papers			Citing papers				
	Date of lead paper	# in cluster	Total citations (mean)	Outdegree (mean)	# in cluster	# cited in the network (%)	Total citations (mean)	Indegree (% total citations)	Outdegree (mean)
C1	2005[25]	103	6320 (61)	2073 (20)	1959	1383 (71%)	68,381 (35)	20,736 (30%)	23,170 (12)
C2	2009[26]	57	4121 (72)	934 (16)	1903	1364 (72%)	55,025 (29)	20,245 (37%)	23,569 (12)
C3	1983[27]	62	3836 (62)	1270 (20)	1265	1006 (80%)	48,105 (38)	17,615 (36%)	20,302 (16)
C4	2000[28]	23	1523 (66)	448 (19)	813	632 (65%)	30,528 (38)	12,345 (25%)	14,552 (18)
C5	2000[29]	12	883 (73)	131 (11)	717	466 (64%)	47,335 (66)	3467 (7%)	4139 (6)
C6	1997[30]	18	627 (52)	227 (13)	591	379 (64%)	19,443 (33)	4869 (25%)	5701 (10)
C7	1980[31]	15	618 (41)	58 (4)	573	365 (64%)	28,028 (49)	2813 (10%)	3204 (6)
C8	1984[32]	27	838 (31)	138 (5)	507	372 (73%)	23,303 (46)	3657 (16%)	2956 (6)
C9	2011[33]	10	650 (65)	24 (2)	475	257 (54%)	20,101 (42)	1485 (7%)	2103 (4)
C10	1996[34]	14	352 (25)	12 (1)	307	184 (60%)	7992 (26)	1222 (15%)	1727 (6)
Others		48	1463 (30)	167 (3)	1402	791 (56%)	64,103 (46)	4961 (8%)	6835 (5)

TABLE 3 Interconnections between clusters. References from (rows) and citations to (columns) each of clusters C1–C10 and all other papers in the network. The yellow highlighted cells display citations within each cluster.

	C1	C2	C3	C4	C5	C6	C7	C8	C9	C10	Others	References
C1	21,600	282	1780	155	353	423	50	428	14	3	155	25,243
C2	620	21,081	339	998	24	553	31	552	30	16	261	24,505
C3	2284	213	17,051	826	61	77	163	643	1	22	231	21,572
C4	437	1245	1273	11,569	20	25	18	181	3	21	208	15,000
C5	420	22	27	4	3671	84	15	2	5	0	20	4270
C6	765	498	155	23	144	4230	0	57	46	1	9	5928
C7	45	11	81	3	12	3	3031	63	0	0	13	3262
C8	226	80	179	31	1	9	84	2467	0	0	17	3094
C9	11	7	5	0	18	58	0	1	2024	0	3	2127
C10	24	73	35	72	0	2	1	29	0	1501	2	1739
Others	223	534	486	164	26	18	27	37	4	4	*2730*	4253
Cites	26,655	24,046	21,411	13,845	4330	5482	3420	4460	2127	1568	3649	

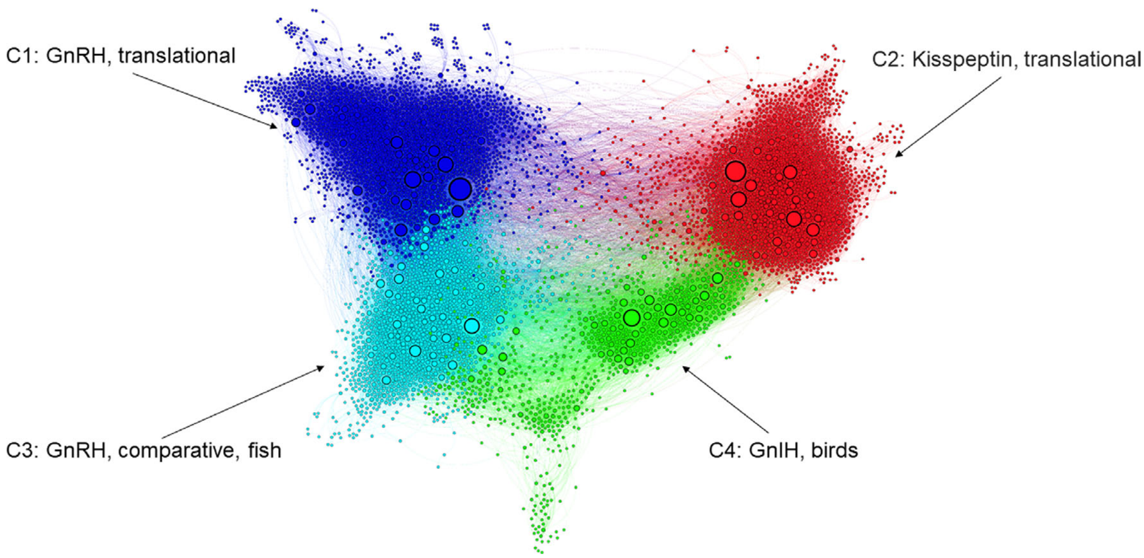

FIGURE 4 Citation network composed of 245 of Bob's papers and 5940 papers that cited these contained in the four largest clusters (C1–C4) described in Figure 3. The node size is relative to indegree, and the node position was determined by the ForceAtlas2 algorithm. Note that the four largest clusters are relatively unaffected by removing the smaller clusters from the network, despite eliminating half of all the papers from the network.

3.2 | Citation network analysis

Figure 3 shows the full citation network, and Table 1 summarises the 25 clusters by common keywords. Nodes in the 10 largest clusters are coloured, while nodes in the remaining 15 smaller clusters are in grey. In what follows, we describe these 10 largest clusters, which include 87% of all papers in the network and 93% of all edges (Table 2). For each cluster, a 'lead paper'—the citing paper with the highest degree in the cluster—is cited in Table 2.

The four largest clusters, C1–C4 are all tightly-connected; their 6185 papers received 217,839 citations, 85,957 (40%) of which came from the network. The other 4716 papers in the network received a similar number of citations (215,736), but only 13% of these came from the network. For each of the clusters C1-C4, between 71% and 96% of within-network references were to other papers in the same cluster and between 77% and 96% of within-network citations were from papers in the same cluster (Table 3). Clusters 5–10 are summarised briefly in the next section, but here we concentrate on C1-C4 as these are each well-defined, highly interconnected networks that each appear to capture a substantial proportion of the core literature on their particular topic. Given our focus, in Figure 4 we visualise a network composed only of C1–C4 after rerunning our layout algorithm. Because the citations to and from these clusters come overwhelmingly from within these clusters or between these clusters, the network structure related to these four clusters remains largely unchanged, but it is now easier to see the links between these clusters.

3.3 | Clusters 1–4

C1: With 103 of Bob's papers and 1959 citing papers, this cluster is concerned with translational application of GnRH agonists and antagonists. Of the 304 citing papers with degree ≥40, 142 had 'GnRH receptor' or an equivalent term in the title, and another 30 had titles that implied studies of GnRH actions on gonadotrophs (including on calcium signalling, second messenger pathways, receptor desensitisation, LH gene expression, and actin cytoskeleton remodelling). The titles of 107 papers indicated that they were on cancers of the ovary, breast, prostate and endometrium; another 35 papers were on expression of GnRH or its receptor at these extra-pituitary sites, but did not mention cancer in the title. Many of these papers addressed how GnRH agonists or antagonists may be therapeutically helpful, and another 18 papers mentioned GnRH agonists and/or antagonists in the title without identifying a particular tissue that was studied. Only 17 of the 304 citing papers studied were included in none of these subclasses: these included 10 reviews with broad titles.

C2: With 57 of Bob's papers and 1903 citing papers this cluster has a translational focus on kisspeptin. Of the 302 citing papers with degree ≥40, 232 had 'kisspeptin' or a directly related term in the title. The others include many that define the context in which kisspeptin neurones act, including papers on the properties of GnRH neurones, on the regulation of luteinising hormone secretion, and on neurokinin (which is co-localised with kisspeptin). Many others define the translational context in which kisspeptin agonists or antagonists may be therapeutically beneficial, including papers on polycystic ovarian syndrome, obesity-associated infertility, and precocious puberty. Not all papers in the network that had kisspeptin in the title belonged to this cluster; 55 were in C4 (below) and 86 (mainly studies in fish), were in smaller clusters outwith the 10 clusters described here. This is a large proportion of the kisspeptin literature: a search on WoS for documents with kisspeptin (or a related term) in the title returned 3745 documents published before 2024, including 62 RPM documents. The whole network includes about 32% of all papers published before 2024 that were explicitly focussed on kisspeptin.

C3: with 62 of Bob's papers and 1265 citing papers, this cluster is focussed on the multiple isoforms of GnRH in different species of fish, their localisation, regulation and possible function. The lead paper announced that fish possess a unique GnRH that differs from mammalian GnRH in two amino acids. Of the 300 papers with a degree ≥40, 177 were on GnRH in fish, as inferred from specific terms in the title, and 59 were on GnRH in other animals, just one of which was in humans. This cluster contains 62 of Bob's papers, but while 18 are on GnRH in fish, 29 are on other animals. A search on WoS for papers with GnRH (or an equivalent term) in the title returned more than 39,000 documents and listed Bob (with 259 documents) as the most prolific contributor. A search within this set for 'fish' as a topic returned 2060 documents, of which Bob was an author of 19. This network captures a substantial proportion of the literature on GnRH in fish, noting that C2 and (to a lesser degree C4) also include papers on this topic.

C4: with 23 of Bob's papers and 813 citing papers this cluster is focussed on GnRH inhibitory hormone (GnIH) in birds and its homologues in different species. Of the citing papers, 338 had 'GnIH' or a related term in the title, and another 301 were identifiable by their title as related to reproductive neuroendocrinology in birds. The remaining papers included 65 reviews that did not specify a species; and research papers on reproduction in diverse species of livestock, fish; mammals including impala, ground-squirrels and bonobos, and other animals including snakes; frogs and mud crabs. Only three papers could be identified by their title as concerning humans. A search on WoS for documents with GnIH (or a related term) in the title returned 916 documents, so it appears that this cluster contains a substantial proportion (∼39%) of the total GnIH literature.

C1 and C3 grew from the earliest of Bob's papers. While the number of citations increased gradually, there was an abrupt increase in the number of citing papers after 2010: in 2000–2010, 266 ± 12 papers cited at least one of Bob's papers compared to 159 ± 7 papers in 1990–1999. This increase is associated with the growth of C2 and C4 after about 2010 (Figure 5). The kisspeptin field, of which C2 is a large subset, grew rapidly with the identification of kisspeptin neurons as regulators of GnRH neuronal activity, and the GnIH field grew rapidly at about the same time. Curiously the abrupt emergence of these new clusters had no visible impact on the year-on-year rate of citation rates to Bob's papers (Figure 1C).

3.4 | Clusters 5–10

C5: 12 of Bob's papers and 717 citing papers. Of the 41 papers with degree ≥30, 24 had 'G protein-coupled receptor' in the title; the others were all on particular G protein coupled receptors—none were specifically on the GnRH receptor.

C6: 18 of Bob's papers and 591 citing papers. Of the 100 citing papers with degree ≥30, 66 had 'hypogonadotropic hypogonadism' or an equivalent term (e.g., Kallman's syndrome) in the title. The others were on specific genetic mutations affecting the GnRH system, mutations of other G-protein receptors, the genetic of reproductive disorders, and therapeutic approaches to these diseases.

C7: 15 of Bob's papers and 573 citing papers. Of the 32 papers with degree ≥30, 30 had 'somatostatin' in the title. Processing of the somatostatin precursor in the brain yields different forms of somatostatin with different lengths, and many papers in this cluster address the biological activity of these different forms.

C8: 27 of Bob's papers and 507 citing papers, most of which were published before 1996. Of the 48 papers with degree ≥30, 41 had GnRH or an equivalent term in the title; 19 of these were explicitly on GnRH gene expression.

C9: 10 of Bob's papers and 475 citing papers. Of the 30 papers with degree ≥20, 13 had prokineticin or a related term in the title; four of these also had vascular endothelial growth factor (VEGF) in the title

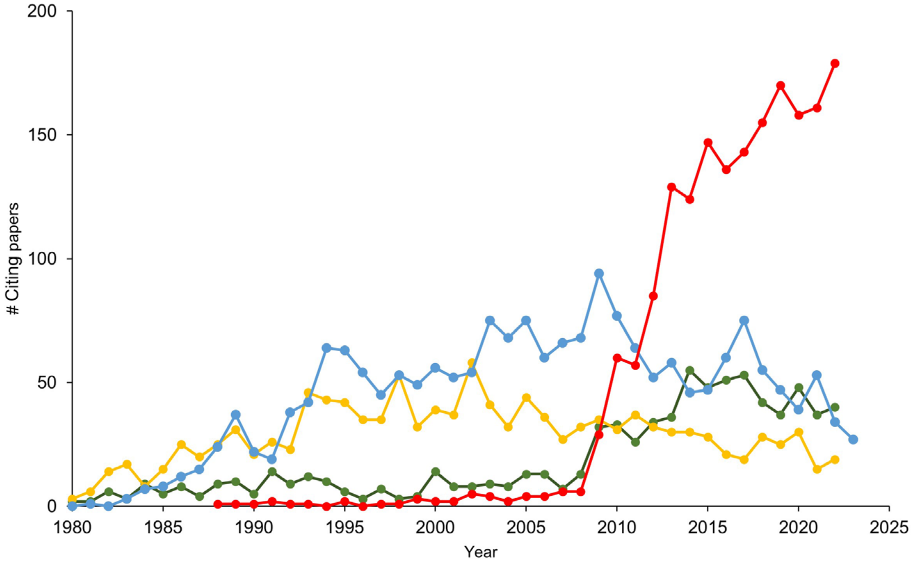

FIGURE 5 Numbers of citing papers per year for clusters C1 (blue), C2 (red) C3 (0range), C4 (green).

and another five had VEGF alone in the title; nine had prostaglandin in the title.

C10: 14 of Bob's papers and 307 citing papers. The lead paper is on 'Reproductive suppression in subordinate, non-breeding female Damaraland mole-rats'.[34] Of the 48 papers with degree ≥20, 40 specify 'mole-rat' in the title and five specify 'spotted hyena'.

3.5 | Citations by cluster

Bob's 245 papers in C1–C4 were cited 15,800 times for an overall average of 64/paper; the 5940 citing papers were cited 202,039 times (34/paper). The size of the difference suggests that Bob's papers may indeed be relatively highly cited by these clusters, but the difference partly reflects the fact that the citing papers, in being published later than Bob's papers, have had less opportunity to be cited.

By contrast, Bob's 96 papers in C5–C10 received 3968 citations (41/paper), whereas the 3170 citing papers in these clusters received 146,202 citations (46/paper). The 48 of Bob's papers in smaller clusters received 1463 citations (30/paper), whereas the 1402 citing papers in these clusters received 64,103 citations (46/paper).

4 | DISCUSSION

We set out to study the oeuvre of one scientist, but we find that the oeuvre is constructed with 925 co-authors. The citations to Bob's papers—coming from more than 10,000 papers—delineate a large and loosely affiliated community with some common interests, mainly concerning aspects of GnRH and its regulation.

The approach that we took in the present analysis is unusual. We constructed a network from the papers of an individual and the papers that cited these, as a way of potentially studying the impact of that individual. We analysed the network by detecting clusters of interconnected papers, and this identified specific foci of research interest.

We evaluated the relative citation impact of Bob's papers in each cluster by comparing them to the papers that cited Bob's papers. We controlled for effects of publication date by comparing the citations to Bob's papers and citing papers that were published in the same time period.

The four largest clusters (C1–C4) were each highly interconnected and appeared to capture a large proportion of the core literature on their particular topic. They defined very different questions: C1 and C2 are *translational*: C1 is concerned with the GnRH receptor and how agonists and antagonists might be therapeutic tools in treating cancers of reproductive tissues; C2 with how kisspeptin neurones behave and how this knowledge can be used to treat conditions associated with dysregulated gonadotrophin secretion. By contrast, C3 and C4 have no overt translational focus. C3 is mainly concerned with the functional roles in fish of the different GnRH isoforms, and C4 with the role of GnIH in avian reproduction.

Bob's 245 papers in clusters C1–C4 were cited (on average) 64 times each compared to 34 times each for the citing papers (Table 1). By contrast, his 144 papers outside these clusters were cited 38 times each compared to 46 times each for the citing papers. We might ask why Bob's papers in clusters C1-C4 are relatively well-cited (i.e., cited more often than the citing papers in the same cluster) when his papers in other clusters are not—what makes C1–C4 different from the other clusters?

The citations that a paper gathers depends on (1) the opportunities for it to be cited –the number of relevant papers that will be published by the community in the next few years multiplied by the number of references in those papers; (2) the utility of the paper to that community—in the sense of how citing it might advance a case made by the citing author; and (3) name recognition by that community of an individual or familiarity with their work. A large, productive community does not in itself mean that its papers are well cited, but some will gain disproportionately many citations. The distribution of citations is very unequal in all areas of science,[35,36] as first characterised by de Solla Price, who recognised that it followed a power law[37] (Figure 2). He proposed that this arose from 'Cumulative Advantage', conceiving that the more often a paper is cited, the more it is noticed; the more it is noticed, the more it is cited, and ultimately the more it is cited, the more it will be cited, as it reaches 'totemic' status, like celebrities who are famous for being famous, even if no-one can quite remember why they were famous in the first place.

The pattern of citations to the oeuvre of an individual is thus not a property of the oeuvre alone, but a construct of the communities that make use of that oeuvre, and is influenced by many factors, including the citation conventions of those communities, their understanding of which journals are most important, and by the perceived authority or esteem of various members of those communities. As Gilbert put it[5] '…an author, in choosing one collection of papers to cite, is not only providing support for his own paper, but is also implicitly displaying his allegiance to a particular section of the scientific community'.

Communities change over time: they expand and dissolve; some research questions are either answered or discarded, while others arise from methodological advances or discoveries, and authors retire from science for diverse reasons. Time also brings changes in the conventions of publishing and citation. In papers that cited Bob, the average number of references doubled between the 1980s and the 2020s, hence, the average number of citations received by a paper published in the present decade will be twice that received by papers published in the 1980s. This trend does not merely inflate the citations of recent papers, but may also increase the inequality of citations.[38] In the 1980s, when all journal articles were published in print, there was pressure on space, and authors were enjoined to only cite papers that were directly relevant. Now, authors can be expansive in their introductions, using their references to establish their knowledge of the field. This means that more papers are cited,[39] which appears democratically virtuous, but this may disproportionately benefit papers that are already well cited, increasing citation inequality.[40] The increased number of references also reflects the increasing prevalence of reviews, and many authors use reviews to find the references that they go on to cite (see[41,42] for problems associated with this); again this factor will increase citation inequality.

Papers in C1–C4 were the source of 86,320 references to other papers in the network and received 85,597 citations in return. This symmetry is to be expected: in the literature as a whole, the total number of citations has to approach the total number of references—and hence the average citation rate will be expected to approach the average size of reference lists in well-connected clusters. In clusters C1–C4, 40% of the citations came from papers in these clusters, other papers in the network (and mostly from papers in the same cluster), and on average, papers in these clusters referred to 14 other papers in the network. C6 shares these features (and again in this cluster Bob's papers are cited more often than the citing papers). However, for the rest of the network, papers are much less cited by other papers in the network, and reference far fewer of them. It appears therefore that a relatively high citation rate comes with being a large part of a compact cluster, with high rates of mutual citation.

To put it another way, a paper that is very relevant to a particular topic and has a particular utility in supporting or extending current understanding, it will be relatively well cited by other papers on that topic. If the 'community' on that topic is small, this will limit how many citations can accrue—and even if citations are unequal within the cluster, the scope for high citations remains limited. It appears that within clusters C1–C4, Bob's papers are well-known and well-cited, even if their citations do not reach extravagant levels.

Conversely if a paper has no focussed 'home' but is tangentially relevant to a large, dispersed community, then it is less likely to gain attention—though if it does attract enough citations then the scope for cumulative advantage may be greater. Thus, when a 'community' is large and dispersed, the citation inequality is likely to be greater—and the mean number of citations will be much higher than the median. It appears that Bob's papers outside C1–C4 have not (yet) become core to any other community, at least to the same extent as those in C1–C4. Of course, C1–C4 contain more of Bob's papers than most of the other clusters, and this perhaps is important in that multiple contributions on a given theme may raise the profile of all those contributions within a community.

What should be apparent from this analysis is that it makes no sense to conceive of the citation impact of an author in isolation; it depends on the existence of a receptive community that is alert to the potential utility of papers from that author, and which uses, amplifies, extends and qualifies the contents of their papers—activities that entail reciprocal referencing and citation between authors. Bob's 389 papers sent out 18,318 references to 9836 papers and received 18,114 citations from 10,512 papers (after removing Bob's self-citations), a reciprocity indicative that many of his papers are salient to a focussed research community.

Bob's contributions appear to be part of the core literature on the topics identified with C1–C4, but it is only the activities of the other members of those communities that make them important. It might be that they judge his contributions to these clusters to have exceptional merit—but merit is a multi-dimensional, qualitative concept that defies naïve quantification and has meaning only in the context of a particular community. Most of the citations to Bob's papers come from papers that are narrowly focussed on particular questions—and which extensively reference and cite each other. In other words, the citation rank of each paper reflects its utility to a particular community; so, insofar as its methodological quality is relevant, it is only its quality *according the general standards of that community* (which may be rigorous or lax), not according to any absolute standards.

Citation metrics measure citations; they do not measure quality in any meaningful sense.[2] Bob's contributions have been in fields that, we think, are strong in analytical rigour, experimental methodology, and intellectual conception. Citation metrics say nothing of these. To judge the quality of Bob's work, you might, indeed, just have to read them.

AUTHOR CONTRIBUTIONS

Gareth Leng: Conceptualization; data curation; formal analysis; investigation; writing – original draft; writing – review and editing. **Rhodri I. Leng:** Conceptualization; data curation; formal analysis; investigation; methodology; software; validation; visualization; writing – original draft; writing – review and editing.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/jne.13430.

DATA AVAILABILITY STATEMENT

The citation network dataset used in this study is openly available in Zenodo at (https://doi.org/10.5281/zenodo.11534257). All bibliographic data included in this study are derived originally from Clarivate™ (Web of Science™) and downloaded in January 2024. © Clarivate 2024. All rights reserved.

ORCID

Gareth Leng https://orcid.org/0000-0002-2388-8466

REFERENCES

1. Ioannidis JP, Boyack K, Wouters PF. Citation metrics: a primer on how (not) to normalize. *PLoS Biol*. 2016;14:e1002542.
2. Asknes D, Langfeldt L, Wouters W. Citations, citation indicators, and research quality: an overview of basic concepts and theories. *SAGE Open*. 2019;9. doi:10.1177/2158244019829575
3. Latour B, Woolgar S. *Laboratory Life: the Construction of Scientific Facts*. Princeton University Press; 1979.
4. Luukkonen T. Why has Latour's theory of citations been ignored by the bibliometric community? Discussion of sociological interpretations of citation analysis. *Scientometrics*. 1997;38:27-37.
5. Gilbert GN. Referencing as persuasion. *Soc Sci Stud*. 1977;7:113-122.
6. Duyx B, Urlings MJE, Swaen GMH, Bouter LM, Zeegers MP. Scientific citations favor positive results: a systematic review and meta-analysis. *J Clin Epidemiol*. 2017;88:92-101.
7. Urlings MJE, Duyx B, Swaen GMH, Bouter LM, Zeegers MP. Citation bias and other determinants of citation in biomedical research: findings from six citation networks. *J Clin Epidemiol*. 2021;132:71-78.
8. Leng G, Lemg RI. *The Matter of Facts: Skepticicm, Persuasion, and Evidence in Science*. MIT Press; 2020.
9. Broad WJ. The publishing game: getting more for less. *Science*. 1981;211:1137-1139.
10. Fanelli D, Laraviere V. Researchers' individual publication rate has not increased in a century. *PLoS One*. 2016;11:e0149504. doi:10.1371/journal.pone.0149504
11. Merton RK. The Matthew effect in science II: cumulative advantage and the symbolism of intellectual property. *ISIS*. 1988;79:607-623.
12. MacRoberts MH, MacRoberts BR. Problems of citation analysis: a critical review. *J Am Soc Inf Sci*. 2014;40:342-349.
13. Oldfield RC. The assessment and analysis of handedness: the Edinburgh inventory. *Neuropsychologia*. 1971;9:97-113.
14. Crespo JA, Li Y, Ruiz-Castillo J. The measurement of the effect on citation inequality of differences in citation practices across scientific fields. *PLoS One*. 2013;8:e58727.
15. Fortunato S, Bergstrom CT, Borner K, et al. Science of science. *Science*. 2018;385:eao0185. doi: 10.1126/science.aoo0185
16. Leng G, Leng RI. Oxytocin: a citation network analysis of 10 000 papers. *J Neuroendocrinol*. 2021;33:e13014.
17. Leng G, Leng RI, Ludwig M. Oxytocin-a social peptide? Deconstructing the evidence. *Philos Trans R Soc Lond Ser B Biol Sci*. 2022;377:20210055.
18. Traag VA, Waltman L, van Eck NJ. From Louvain to Leiden: guaranteeing well-connected communities. *Sci Rep*. 2019;9:5233.
19. Jacomy M, Venturini T, Heymann S, Bastian M. ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software. *PLoS One*. 2014;9:e98679.
20. Millar RP, Glover TD. Seasonal changes in the reproductive tract of the male rock hyrax, Procavia capensis. *J Reprod Fertil*. 1970;23:497-499.
21. Singh A, Lal B, Kumar P, Parhar IS, Millar RP. Role of neurokinin B in gametogenesis and steroidogenesis of freshwater catfish, *Clarias batrachus*. *Cell Tissue Res*. 2023;393:377-391.
22. Parolo PDB, Pan RK, Ghosh R, Huberman BA, Kaski K, Fortunato S. Attention decay in science. *J Informet*. 2015;9:734-745.
23. Larivière V, Archambault E, Gingras Y. Long-term variations in the aging of scientific literature: from exponential growth to steady-state science (1900–2004). *J Am Soc Inf Sci Technol*. 2007;59:288-296.
24. de Solla Price DJ. Networks of scientific papers. *Science*. 1965;149:510-515.
25. Cheng CK, Leung PC. Molecular biology of gonadotropin-releasing hormone (GnRH)-I, GnRH-II, and their receptors in humans. *Endocr Rev*. 2005;26:283-306.
26. Oakley AE, Clifton DK, Steiner RA. Kisspeptin signaling in the brain. *Endocr Rev*. 2009;30:713-743.
27. Sherwood N, Eiden L, Brownstein M, Spiess J, Rivier J, Vale W. Characterization of a teleost gonadotropin-releasing hormone. *Proc Natl Acad Sci USA*. 1983;80:2794-2798.
28. Tsutsui K, Saigoh E, Ukena K, et al. A novel avian hypothalamic peptide inhibiting gonadotropin release. *Biochem Biophys Res Commun*. 2000;275:661-667.
29. Sealfon SC, Chi L, Ebersole BJ, et al. Related contribution of specific helix 2 and 7 residues to conformational activation of the serotonin 5-HT2A receptor. *J Biol Chem*. 1995;270:16683-16688.
30. Pradayrol L, Jornvall H, Mutt V, Ribet A. N-terminally extended somatostatin: the primary structure of somatostatin-28. *FEBS Lett*. 1980;109:55-58.
31. de Roux N, Young J, Misrahi M, et al. A family with hypogonadotropic hypogonadism and mutations in the gonadotropin-releasing hormone receptor. *N Engl J Med*. 1997;337:1597-1602.
32. Seeburg PH, Adelman JP. Characterization of cDNA for precursor of human luteinizing hormone releasing hormone. *Nature*. 1984;311:666-668.
33. Martin C, Balasubramanian R, Dwyer AA, et al. The role of the prokineticin 2 pathway in human reproduction: evidence from the study of human and murine gene mutations. *Endocr Rev*. 2011;32:225-246.
34. Bennett NC, Faulkes CG, Molteno AJ. Reproductive suppression in subordinate, non-breeding female Damaraland mole-rats: two components to a lifetime of socially induced infertility. *Proc Biol Sci*. 1996;263:1599-1603.
35. Bornmann L, Leydesdorff L. Skewness of citation impact data and covariates of citation distributions: a large-scale empirical analysis based on Web of Science data. *J Informet*. 2017;11:164-175.
36. Ruiz-Castillo J, Costas R. The skewness of scientific productivity. *J Informet*. 2014;8:917-934.
37. de Solla Price D. A general theory of bibliometric and other cumulative advantage processes. *J Am Soc Inf Sci*. 1976;27:292-306.
38. Brzezinski M. Power laws in citation distributions: evidence from Scopus. *Scientometrics*. 2014;103:213-228.
39. Larivière V, Gingras Y. The decline in the concentration of citations, 1900–2007. *J Am Soc Inf Sci Technol*. 2009;60:858-862.

40. Varga A. The narrowing of literature use and the restricted mobility of papers in the sciences. *Proc Natl Acad Sci USA*. 2022;119: e2117488119.
41. Leng RI. A network analysis of the propagation of evidence regarding the effectiveness of fat-controlled diets in the secondary prevention of coronary heart disease (CHD): selective citation in reviews. *PLoS One*. 2018;13:e0197716.
42. Greenberg SA. How citation distortions create unfounded authority: analysis of a citation network. *BMJ*. 2009;339:b2680.

How to cite this article: Leng RI, Leng G. A career in numbers: A citation network analysis of the work of RP Millar and his contribution to GnRH research. *J Neuroendocrinol*. 2024; e13430. doi:10.1111/jne.13430