How do life sciences cite social sciences? Characterizing the volume and trajectory of citations

Reference:
Zhou Hongyu, Sun Beibei, Guns Raf, Engels Tim, Huang Yin, Zhang Lin.- How do life sciences cite social sciences? Characterizing the volume and trajectory of citations
Journal of the Association for Information Science and Technology / Association for Information Science and Technology - ISSN 2330-1635 - (2024), p. 1-16
Full text (Publisher's DOI): https://doi.org/10.1002/ASI.24891

How do life sciences cite social sciences?

Characterizing the volume and trajectory of citations

Hongyu Zhou

Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp (Belgium)

Beibei Sun

Center for Science, Technology & Education Assessment (CSTEA), Wuhan University, Wuhan (China)
School of Information Management, Wuhan University, Wuhan (China)

Raf Guns

Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp (Belgium)

Tim C.E. Engels

Centre for R&D Monitoring (ECOOM), Faculty of Social Sciences, University of Antwerp, Antwerp (Belgium)

Ying Huang

Center for Science, Technology & Education Assessment (CSTEA), Wuhan University, Wuhan (China)
School of Information Management, Wuhan University, Wuhan (China)
Centre for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Leuven (Belgium)

Lin Zhang*

Center for Science, Technology & Education Assessment (CSTEA), Wuhan University, Wuhan (China)
School of Information Management, Wuhan University, Wuhan (China)
Centre for R&D Monitoring (ECOOM) and Department of MSI, KU Leuven, Leuven (Belgium)

*: **Correspondence concerning this article should be addressed to Prof. Lin Zhang,**

Email: zhanglin_1117@126.com

ABSTRACT

Social sciences are increasingly recognized as significant for building a sustainable world since the social perspective can assist researchers in other fields in navigating public controversy and designing more responsible interaction mechanisms between the natural and social systems. However, the question arises: to what extent do natural sciences rely on social science research in their studies? Examining life science publications from seven PLoS journals, this paper attempts to characterize the volume and trajectory of citations from life sciences to social sciences. We explore three core questions: To what extent do life sciences cite social sciences? What actors in the life sciences are citing social sciences? Which actors in the social sciences are being cited? Our analysis estimates social sciences influence 15–19% of life science publications, contributing to 1.1–1.5% of references in 2018. Social science citers are found across peripheral and central topics of life science disciplines. Cited social science publications exhibit various levels of interdisciplinarity and achieve the greatest citation impact among peers. Citations to social sciences are prevalent in both theoretically and methodologically oriented sections. We show empirically the increasing impact of social sciences on the development of the life sciences.

1 INTRODUCTION

Since Comte's classification of science, the divide and interaction between natural sciences and social sciences have remained a focal question in both the qualitative (Kuhn, 2012) and quantitative (Price, 1970) understandings of science. The two major branches of science differ significantly in research subject, research culture (Kagan, 2009), research strategy (Jaffe, 2014), epistemology (Bonaccorsi, 2022), publication channel (Larivière, Archambault, et al., 2006), researcher productivity (Neumann, 1977), research collaboration (Larivière, Gingras, et al., 2006; Stefaniak, 2001), reference literature (Glänzel & Schoepflin, 1999), the language used (Huang & Chang, 2008; Kulczycki et al., 2020), funding ratios (Xu et al., 2015), etc.

Despite the great cognitive and practical discrepancy, the intellectual exchange between the two is also intensively discussed, arguably more on social sciences benefiting from natural sciences. For instance, the term "Social Physics", nowadays Sociology, was coined by Comte with the ideal of discovering universal laws of human society in a similar fashion to physics. Some claimed that social sciences, at an early age, attempted to emulate the natural science ideal, which gives rise to the cumulative development of social science knowledge, for example, the mathematization of economics. A similar trend can also be found in Sociology. In 1946, 54% of substantive articles in the *American Sociological Review* (ASR) were completely theoretical without any form of mathematical analysis, but that number dropped to 12% in 1976 (Heise & Simmons, 1985). Such affinity did not fade with time, yet became more prominent with the rise of digital scholarship, big data, and biotechnology. The recent zeitgeist in computational social science seems to empower many social science disciplines with natural-science-like research technologies to conduct quantitative studies using human behavioural big data. Brain-imaging, such as PET and fMRI, has also gained vast attention and usage in the study of *economics* (Camerer et al., 2004), *psycholinguistics* (Gernsbacher & Kaschak, 2003), *information science* (Gwizdka et al., 2013), *media studies* (Anderson et al., 2006), and *sociology* (Pickersgill, 2013), by visually recording various forms of neural activity in response to certain interventions in social contexts. These research technologies borrowed from natural sciences are believed to be significant ingredients to obtaining a high-consensus rapid discovery model as observed in natural sciences (Collins, 1994).

On the other side of the coin, natural sciences learning from social sciences is also repetitively encouraged in philosophy, widely practiced in education, yet poorly evidenced in the actual production of science. As early as 1958, Harvard University initiated a fellowship program to provide rigorous training in social sciences to younger medical professionals and researchers (Medical and Social, 1958). As for college education, courses from social sciences can be found in the curriculum of many natural science disciplines with social implications, such as behavioral science and psychology in medical science, and economics in engineering, which leads to increasing involvement of social sciences in academic degrees in STEM fields. In January 2021, the U.S. Department of Homeland Security updated the list of STEM-designated degree programs with 8 new programs with connections to social sciences, such as human-centered technology design and mathematical economics.

Besides a growing presence in education, many natural scientists also celebrated the pre-eminence of social science knowledge in the scientific research of human-influenced or human-dominated systems. Meloni (2014) argued that "biology is becoming social" by illustrating how evolutionary biology, neuroscience, and molecular biology are adjusting their view take on altruism, brain, and gene, respectively, to a more social perspective. Conservation biology is perhaps one of the most active disciplines in disseminating social science knowledge to their domain researchers through literature (e.g., Moon & Blackman, 2014). There are not only calls to understand social science knowledge and philosophy but also to mainstream social sciences in conservation research to enable "ecologically effective and socially just conservation". Bennett et al. (2017) introduced a full spectrum of conservation social science, with 18 interdisciplinary common grounds between the two, to facilitate a better understanding of the potential of integrating social science knowledge in conservation research. In synthetic biology, an emerging field that can contribute both rice producing beta-carotene and bioweapons, social scientists are invited to take a more vital role in research programs to provide discussions on ethical, legal, and social implications (ELSI), a required proponent in government-funded research from both Europe, UK, and USA (Calvert & Martin, 2009). However, it is often expected for social scientists to be "contributors", who contribute ELSI ex post facto, or 'brokers' between scientists and the public, helping scientists navigate public controversy (Calvert & Martin, 2009).

Based on the above-discussed examples, we hypothesize that natural science research mostly solicits assistance from social sciences after the research has been carried out (in discussion of implications), or before as a general philosophical guide (encourage learning the social perspective), but scantly in the course of research (for methods and relevant knowledge). That is to say, social science knowledge may be learned from, remembered, and invited, but less frequently cited in the natural science literature. Can one trace, pinpoint, and characterize the knowledge diffusion process from social sciences to natural sciences in the scientific literature? This constitutes our overall research question.

In this paper, we take life sciences, a branch of natural science studying living organisms and life processes, as an example and hope to contribute in-depth and holistic explorations of the knowledge flow from social sciences to natural sciences. We tackle this topic by asking three sets of questions as follows:

- ***To what extent do life sciences cite social sciences?*** What is the degree of such citations and how does it evolve over time? Do life sciences cite social science knowledge? Do they obtain this knowledge directly from social science venues/communities? Do they obtain social science knowledge directly from social scientists?
- ***What kind of research in life sciences is citing social sciences?*** Does research in life sciences that employs knowledge from social sciences only work on peripheral or interdisciplinary topics?
- ***What are the characteristics of the cited social sciences knowledge? And for what purpose?*** Does research in life sciences cite social sciences with high-impact? During such interdisciplinary knowledge seeking, does the cited social science research tend to

be more disciplinary or interdisciplinary? Are social sciences cited in sections that primarily provide theoretical or contextual discussions or in those that focus on methods or research design?

In the following sections, we first introduce the data we employed and the analyses conducted for this study. We then present and discuss the results in section 3 and the last section concludes.

2 DATA AND METHODOLOGY

We start by introducing our selected datasets and respective processing procedures. In section 2.2, we discuss the research questions and explain our conceptualization and operationalization of these questions. Figure 1 gives a general outline of this study.

2.1 Data

2.1.1 PubMed Central

To construct a viable dataset on life science publications, we chose PubMed Central and obtained 230,585 publications records and their full-text data from PubMed Central as the subject of this study. We believe it not only provides a good representation of many important branches of life science, such as biology and medicine, but also curates detailed bibliographic information and full-text data which can help inform both the volume and location of references. Our dataset contains articles published from 2005 to 2018 in seven journals, namely, *PLoS Biology* (n=2,417), *PLoS Computational Biology* (n=5,486), *PLoS Genetics* (n=6,970), *PLoS Medicine* (n=1,611), *PLoS Neglected Tropical Disease* (n=5,735), *PLoS One* (n=202,389), *and PLoS Pathogens* (n=5,977). From the obtained full-text data (in XML format), we followed the procedures described by Thijs (2020) and retrieved three types of information:

- Bibliographic information such as title, journal, publication year, and DOI;
- Bibliographic information for the references listed in the article, for instance, title, author, journal, sequence, etc.;
- Section structures/names and corresponding references (sequence in the reference list) are mentioned within.

To obtain unified section structures for all publications, we mapped the section names into one of the five categories: **I**ntroduction, **M**ethodology, **R**esults, **D**iscussion, and Others (IMRaD). The details of mappings are described in the Appendix A1.

2.1.2 Microsoft Academic Graph

A second dataset is harvested from Microsoft Academic Graph (MAG) to provide the additional necessary information, namely article-level classification and general bibliometric information.

First, we used the article-level disciplinary classifications from MAG to assign publications and references in our dataset to disciplines. Since the first dataset does not provide complete DOIs for references, we, therefore, matched references from PubMed with publications in MAG using bibliographic information such as title, author name, journal, and year. In total, 10,045,358 references were successfully recognized and classified (95.2% of the total). We used the second level of MAG's field classification systems, which recognizes more than 290 fields. This classification serves two tasks: identifying social science references, and identifying life science publications. In the first task, 60 fields are selected to represent social sciences following the approach from our previous study (Zhou et al., 2022). The second task is conducted to address the multidisciplinarity of *PLoS One* by identifying life science publications within its portfolio. Since life sciences are not provided as a category in MAG classification, or as an individual branch in other major bibliographic databases, we selected 75 fields from three areas of the life sciences (see details in Appendix A3), namely, biological and environmental sciences (22 fields), medical and health sciences (43 fields), and agricultural, earth, and marine sciences (10 fields). Focusing on these fields, we retained 169,895 life science publications in *PLoS One* (83.9%). For the other journals, we kept all their publications since they are discipline-specific journals with a focus on certain topics in life sciences.

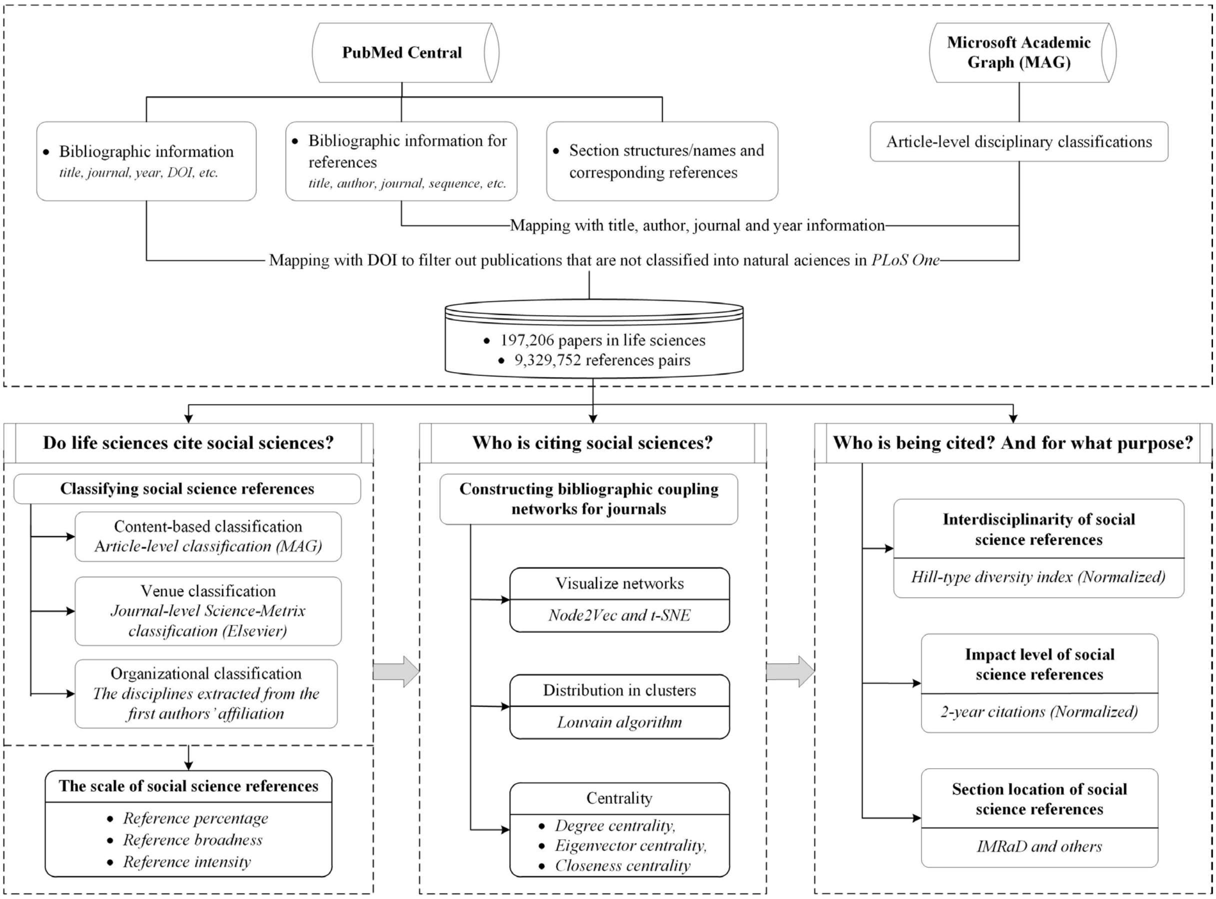

Figure 1. Diagram demonstrating the research design of this study. The upper panel introduces the used dataset and processing procedures, and the lower panel describes the three research questions and three corresponding analyses.

In addition, after having publications and their references identified in MAG, we further retrieved several bibliometric indicators, for instance, the number of citations after two years and the level of interdisciplinarity, to assist our analyses. We introduce details of these indicators in the next section.

2.2 Analyses

In the Introduction section, we highlighted three groups of questions that discuss the relationships between life sciences and social sciences. Accordingly, we designed three groups of analyses to tackle these questions one by one.

2.2.1 To what extent do life sciences cite social sciences?

For the first group of questions, we aim to quantify the scale of social science references in our recognized life science publication set. For each reference, we extracted three types of information on their disciplinary origin, namely, the discipline of the published contents (content-based classification), the discipline of the publication venue (venue classification), and the discipline of the first author's affiliation (organizational classification). The three classification schemes were adopted accordingly to discuss three questions: Do life sciences cite social science knowledge (content-based classification)? Do they obtain this knowledge directly from social science venues/communities (venue classification)? Do they obtain social science knowledge directly from social scientists (organizational classification)? We believe the employment of three different classification schemes not only contributes a nuanced understanding of the impact of various actors in social sciences on life sciences but also can be seen as robustness checks for the research question. Regarding operationalization, the discipline of the published contents was assigned using the article-level classification from MAG, which is based on their textual contents (Wang et al., 2020). This classification indicates the origin of the cited knowledge. The disciplines of the publication venues were assigned using the journal-level Science-Metrix classification scheme (Archambault et al., 2011), indicating the community in which the cited knowledge is embedded. The disciplines of the first authors' affiliations were obtained and processed from the affiliation texts recorded in MAG. We extracted the departmental affiliations (secondary institutions such as department, school, or center) of the first author and mapped them into disciplines using an improved thesaurus from our previous studies (Zhang et al., 2018). A detailed introduction to extracting disciplines from affiliation texts can be found in Appendix A2.

To quantify the scale of social science references, we selected three indicators. We show in our previous study (Zhou et al., 2023) that the latter two variables capture two distinct aspects of the first variable, and all three together provide interesting indications of interdisciplinary knowledge flow.

- *Reference percentage* - the percentage of references from social sciences in all references. It answers the question "How much weight do social sciences account for in the entire knowledge base of life sciences";
- *Reference broadness* - the percentage of papers that cite at least one reference from social sciences. It signals the breadth of influence of social science reference;

- *Reference intensity* - for papers that cite references from social sciences at least once, the percentage of social science references in all references. It describes the intensity of the influence of social science reference.

We conduct the analyses with both including and excluding Psychology from social sciences and examine the scale of social science references with or without Psychology. We consider the inclusion/exclusion of Psychology from social sciences for two reasons. First, given that our publications focus primarily on biological and life sciences, we speculate Psychology plays an important role in their research. We are therefore curious to know if life science publications cite other social science disciplines as well. Second, it is debatable to classify references from some subfields or interdisciplinary topics in Psychology as social sciences, for instance, neuropsychology or biological psychology. To achieve a more robust estimate, we believe it is necessary to conduct the analyses both with and without including Psychology.

2.2.2 What kind of research is citing social sciences?

The second group of analyses investigates the characteristics of life science publications citing social sciences (which we will hereafter refer to as 'social science citers'), if any. Specifically, we focus on the cognitive location of social science citers.

Does life science research that cites social sciences mainly come from peripheral regions or interdisciplinary topics of their discipline? Or do social science publications also exert influence on the central or core topics in life science disciplines? To answer this question, we first constructed bibliographic coupling networks for each of the seven journals in our dataset, with publications as nodes and the number of shared references between them as edge weight. We used these networks to obtain a cognitive map of the discipline(s) in which the journal operates. The networks were processed using Node2Vec (Grover & Leskovec, 2016) to obtain their meaningful structures (50-dimensional vector representation) and then fed to t-SNE (Van der Maaten & Hinton, 2008) for dimension reduction (2-dimensional representation) and visualization, following procedures reported by Shen et al. (2019). We visualized the bibliographic coupling networks for each journal and highlighted the locations of social science citers. Next, we applied the Louvain algorithm on the obtained networks to identify clusters that denote different research topics in journals. This allows us to investigate if the practice of citing social sciences is limited to only a few specific topics or occurs more widely and evenly in various topics in journals. In addition, we also quantify the level of the cognitive periphery of social science citers by calculating several node centrality measures in the bibliographic networks of journals, namely degree centrality, closeness centrality, and eigenvector centrality. Since it is computationally intensive to calculate several node centrality measures for large networks (~200k nodes), we take only publications from 2018 (~12k) to represent the journal *PLoS One*.

2.2.3 What are the characteristics of cited social sciences knowledge? And for what purpose?

The third group of analyses focuses on the characteristics of cited social science references and hopes to understand what kind of social science research is cited by life sciences and for what purpose.

We focused on two characteristics of the cited social science publications, namely, interdisciplinarity and citation impact. The practice of citing social sciences can be deemed an interdisciplinary knowledge-seeking behaviour. Does life science research need mostly disciplinary knowledge from cited social science disciplines or does it rely more on interdisciplinary publications that study the same problems from a different lens? To examine the level of interdisciplinarity of cited references, we employed a Hill-type "true diversity" index (Hill, 1973) to quantify interdisciplinarity as the diversity of disciplinary categories in references. The index is designed as follows:

$$\left(\sum_{i=1}^{N} p_i^2\right)^{-1}$$

where *N* denotes the total number of categories and p_i denotes the proportion of references in category i. Consider two papers A and B: A contains references from two disciplines, each accounting for 50%, and its diversity under this index is 2. B contains references from five disciplines, each contributing 20% references. Hence, its Hill diversity index is 5. B, therefore, can be regarded as more interdisciplinary than A since it utilizes more diverse knowledge across disciplines.

Since disciplines exhibit different levels of interdisciplinarity (Zhou et al., 2022), we normalized the interdisciplinarity values of cited publications as the corresponding percentile among publications published in the same field in the same year. In this way, we ensured that we could assess the relative level of interdisciplinarity of social science references without being confounded by disciplinary and temporal dynamics.

In addition, we were also curious to know whether the cited publications are associated with greater citation impact. We obtained the citation counts for all the social science references and normalized them as the citation percentile among publications with the same publication type, from the same field, and in the same year. We presented the distribution of interdisciplinarity percentile and citation impact of cited social science publications and examined their relative status compared with their disciplinary and coeval peers.

The analysis concludes by analyzing the section distribution of social science references from our publication set. Previous studies have shown that the sections where citations are embedded are indicative of their general function since such logical writing structures illustrate the interpretation of cited articles by citing authors (Bertin & Atanassova, 2014; Tahamtan & Bornmann, 2019).

Put together, we hope to contribute to the literature a nuanced understanding of the nexus between life sciences and social sciences and shed light on the connotation and dynamics of interdisciplinarity.

3 RESULTS AND DISCUSSION

3.1 The prevalence of social science references in life sciences

We begin the analyses with a thorough estimate of the prevalence of citations from life sciences to social sciences and explore the following questions: Does research in life sciences cite social science knowledge? Do they retrieve knowledge directly from social science venues and researchers or through immediate knowledge translators?

As shown in Figure 2 (a), we find that research in life sciences not only cited social science knowledge (content-based classification) but also directly cited their communities (venue classification) and researchers (organizational classification). In 2018, 14.7% of publications in our publication set for life sciences cited at least one reference from social sciences under content-based classification on references. Social science knowledge accounts for 1.5% of references in all publications (reference percentage) and 10.1% of references for publications that cite social sciences (reference intensity). On the other hand, looking at citations to social science journals, the broadness of influence is slightly greater than that of citing social science knowledge, yielding 1-2% more citing publications since 2012. The percentage and intensity of references from social science journals were around 1.5% and 9.1%, respectively, in 2018, roughly similar to that of social science knowledge. As for citations to publications with first authors from social science departments, we observed that 19.1% of life science publications in our data cited at least one reference from this category in 2018. However, references in this category only occupy 1.1% of all references and are associated with 5.7% reference intensity on average in 2018. Put together, we estimate that social sciences influence around 15 to 19% of life science publications and account for 1.1 to 1.5% of references in 2018. According to the strictest interpretation (a reference belongs to social sciences only if all three classifications agree), we find that 7.8% of life science publications in 2018 are classified as citing social sciences. According to the broadest interpretation (a publication belongs to social sciences if one classification classifies it as such), 28% of publications cite social science references.

Another pattern we observe is the growing prevalence of social science references in our life science publication set under all three classification methods and all the employed measures. For instance, under the content-based classification of references, the broadness of social science references grew from 12.1% in 2005 to 14.7% in 2018. The percentage of social science references also increased from less than 1% to 1.5% over 13 years. Looking at journals separately, we find that some journals do not exhibit growing usage of social science references and only maintain them at a lower level. Throughout the observed period in our dataset, only 1%-2% of publications in *PLoS Pathogens* and 4%-7% of publications in *PLoS Genetics* cited social science references and we could not find upward trends in usage. It is therefore important to note that although we regard life sciences as an integral set in this analysis, one could observe different patterns in some subfields in life sciences that are weakly influenced by social sciences. Our estimations and observations cannot be lifted out of their embedded contexts.

On the other hand, it should be further noted that Psychology contributes the majority of social science references in our life science publication set. Excluding Psychology from these analyses, one can see that the influence of social sciences without Psychology is substantially smaller, although not negligible. In Figure 2 (b) we show that the broadness of social knowledge excluding Psychology was 8.5 % in 2018, down by 42.2% compared to including Psychology. The shrink in reference percentage is more significant, which decreased to 0.35% under content-based classification, 0.40% under venue classification, and 0.55% under organizational classification (shrink by 76%, 72%, and 48%, respectively). Nonetheless, we could also find the growing prevalence of social science references in recent years, as shown by the increasing value for three measures and all three classification schemes. In summary, we estimate that social sciences excluding Psychology influence around 8–14% of life science publications and account for 0.4–0.6% of references in 2018.

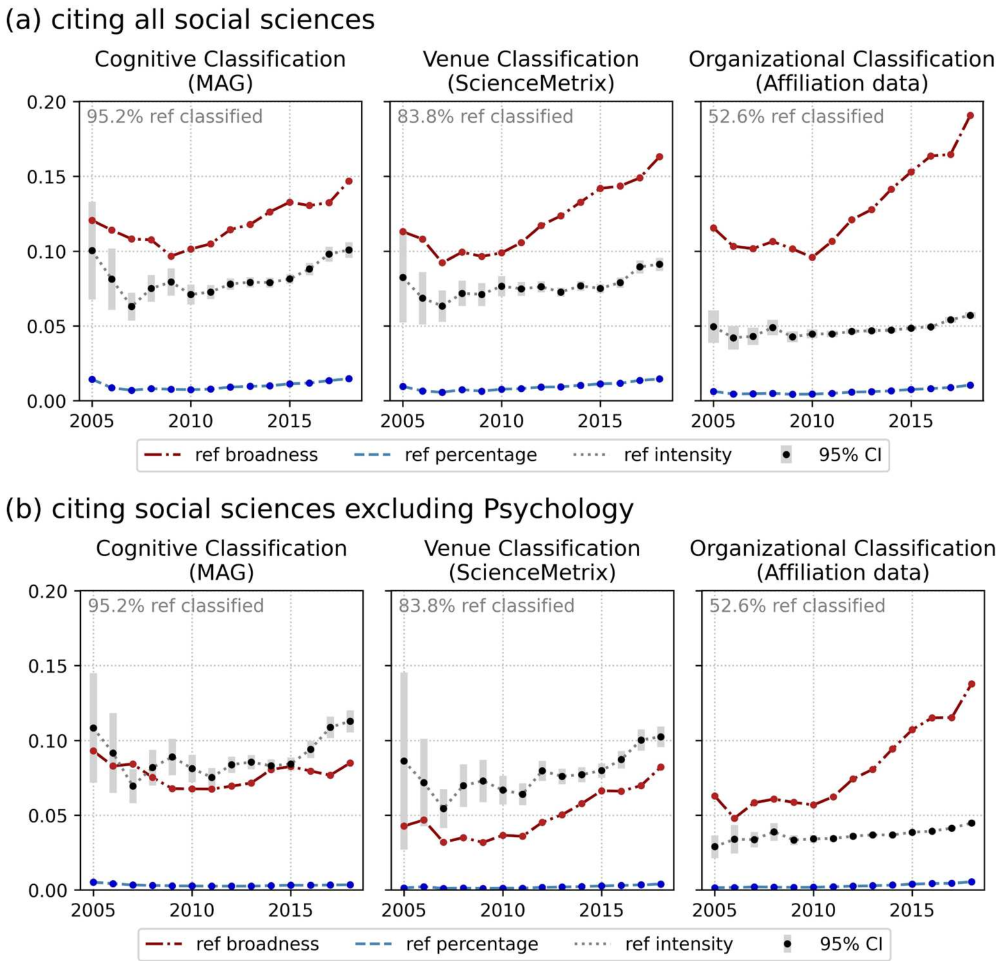

Figure 2. The growing prevalence of social science references in life sciences.

3.2 The cognitive location of social science citers

After confirming and estimating the presence of social science references in our life science publication set, we further question the characteristics of these social science citers in life sciences.

We begin by visually displaying the obtained cognitive maps of six discipline-specialized journals and one multidisciplinary journal (*PLoS One*) in our dataset and pinpoint the relative location of publications that cite at least one reference from social science journals, as shown in Figure 3. We observe that publications that cite social sciences are widely distributed in the cognitive terrain of respective journals and appear in multiple clusters (topics). For instance, looking at the 17 largest clusters in each journal (covering around 89% to 100% of publications in discipline-specific journals and 69% for *PLoS One*), we find that social science citers exist in the majority of clusters in journals, ranging from 12 for *PLoS Biology* to 17 for *PLoS Neglected Tropical Diseases*. If including Psychology in social sciences, all 17 largest clusters in each journal include publications that cite social sciences. In addition, we observe that social science citers can not only be found in the peripheral topics of each journal but also in central regions, as indicated by the distribution of centrality of these publications in the respective journals (Figure 4c).

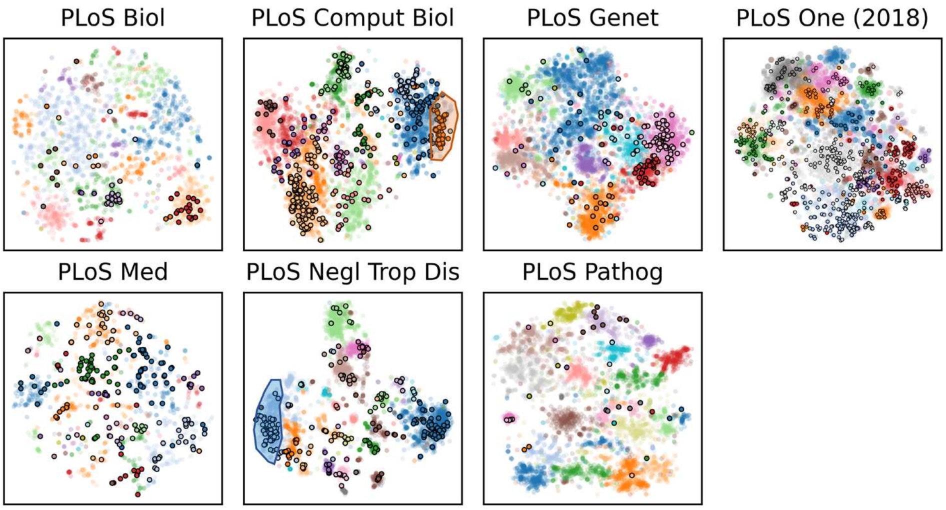

Figure 3. The cognitive location of social science citers in seven journals (using venue classification and excluding psychology). Each dot represents a publication and publications that cite at least one reference from social science journals are highlighted with solid black circles. Different colors denote clusters. The 20 biggest clusters for *PLoS One* are colored. The two polygons indicate selected clusters that are discussed in the main text.

However, the level of concentration of social science citers across clusters is different by journals. Through visual inspection, we can see from Figure 3 that two clusters in the lower region of *PLoS Biology* are populated with social science citers but they are rare in the other clusters. On the other hand, for *PLoS Medicine*, it is difficult to visually determine which cluster hosts more social science citers. To provide a precise assessment of the level of concentration, we calculated the Gini coefficient of the number of social science citers across the 17 largest clusters for each journal and presented their distribution through Lorenz-curve-style plots in Figure 4a. *PLoS Genetics* and *PLoS Biology* exhibit the highest level of concentration of social science citers in topics, with a Gini coefficient of 0.62 and 0.57, respectively. In these two journals, around 60% of the social science citers can be found in only two clusters. The two clusters in *PLoS Biology* relate to research on topics from *Ecology* such as *biodiversity* and topics from *Neuroscience* such as *visual cortex and perception*. The two clusters in *PLoS Genetics* relate to research on topics such as *population genetic structure* and *natural selection*. Since these topics are deeply embedded in sociological, psychological, or demographical contexts, the participation of social sciences is essential in their research agenda. However, social science citers in *PLoS Biology* are on average associated with lower centrality percentiles (eigenvector and closeness), while social science citers of *PLoS Genetics* are mainly found in the higher-centrality regions of the network. This signals different coupling relationships between the life sciences and the social sciences that could emerge in the peripheral topics as well as the core topics in disciplines. On the other hand, in some other journals (for example, *PLoS Pathogens*) social science citers are more evenly distributed across clusters. In these cases, social science references are utilized sporadically and solely for the citing publication itself to serve their special research setting, not as a collective demand for their research topic in general.

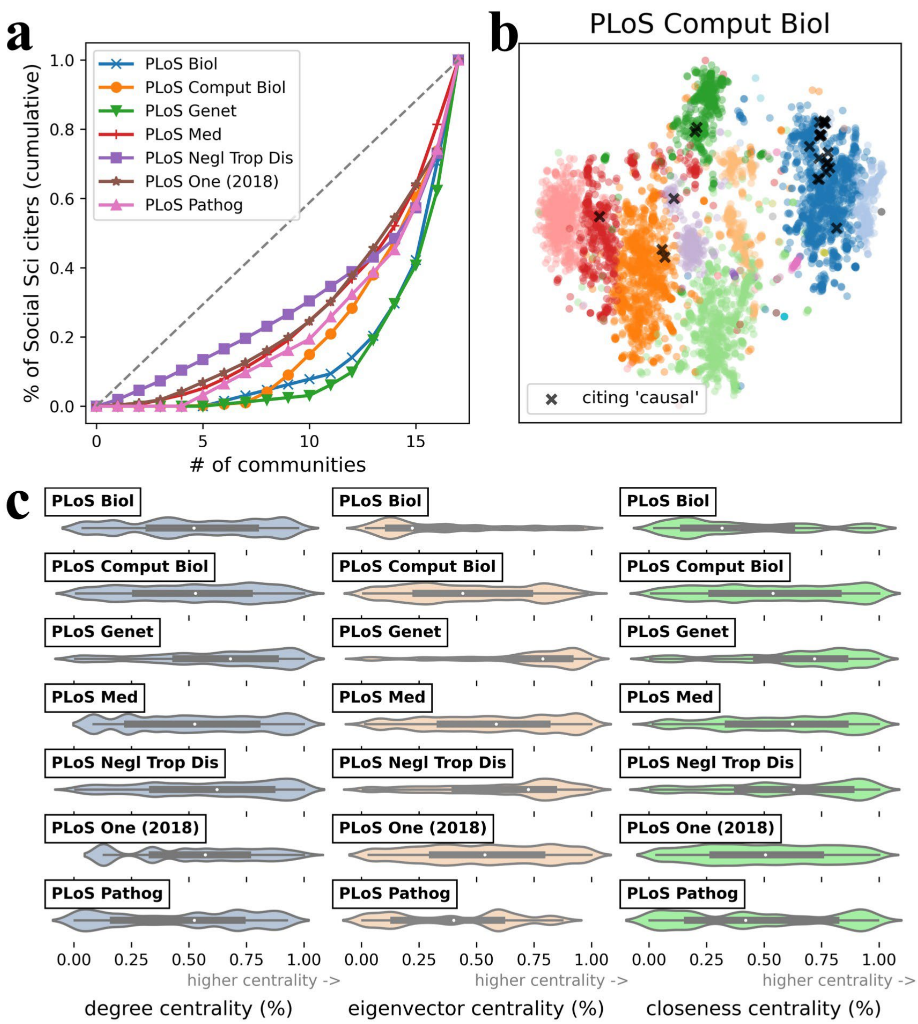

Figure 4. The concentration and centrality of social science citers. (a) the concentration of social science citers in communities; (b) case study: publications citing references that contain "causal" in titles, highlighted as stars in the plot; (c) the percentile of centrality (%) of social science citers.

We also observe some knowledge combinations that bridge concepts from life sciences and social sciences and form interdisciplinary topics. We discuss two examples. The first example (the orange polygon in Figure 3) involves the highlighted cluster in *PLoS Computational Biology* in Figure 3. Upon examination of the titles of the publications that cite social sciences, we find that the most frequent keywords of these publications are *phasic dopamine, dopamine* *signals, decision-making, and percentual decision*. The most cited social science keywords retrieved from references' titles are *prospect theory, risk, choice,* and *decision*. These publications feature research that combines economic theories and computational models to understand decision-making behaviors under risk. For instance, in a paper by Chen et al. (2017) on explorative motor learning, they adopted the relative understanding of gain and loss in prospect theory (Kai-Ineman & Tversky, 1979), i.e. interpret gain or loss relative to a certain reference point, and studied the "action change relative to the current highest score" (p.16). Building on prospect theory, they suggested modeling explorative motor learning as "a sequential decision-making process that is adjusted for motor noise" (p.1). In the second example (the blue polygon in Figure 3 for *PLoS Neglected Tropical disease*), we observe life science publications on parasitic infections such as *Helminth* or *Schistosoma mansoni* are citing social science publications with keywords such as *school children, educational enrollments,* and *wealth effects.* One frequently cited publication in this cluster is from *Demography*, titled "*Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India*" (Filmer & Pritchett, 2001). 70% of citations from this journal to this paper are found in this cluster. This paper proposed proxy indicators of household wealth based on the asset ownership of individuals. Research on these infections, therefore, utilized the proposed indicators to assess the wealth status of infected individuals to uncover the association between socio-economic status and infections (Campbell et al., 2017; Knopp et al., 2010), which contributes to effective prevention or treatment strategies.

In addition to topic-specific knowledge combinations, we also found some social science knowledge that receives attention from multiple topics within journals, for instance, interest in causal inference from *Econometrics*. Figure 4b shows that publications citing references relating to causal inferences span multiple clusters and are widely distributed in the cognitive map of *PLoS Computational Biology*. However, we would also like to point out that the application of causal inference in this field seems to strike root in biological research and emerged variants of causal models that are more applicable to biology research, for example, Dynamic Causal Modelling for fMRI (Stephan et al., 2008), granger causality for attention networks (Wen et al., 2012). The cited publication on causal inference is predominantly published in Biology journals such as *Neuroimage* or *Journal of Neuroscience*, instead of journals related to *Econometrics*. In these cases, it is debatable to classify these references as social sciences, at least not directly.

3.3 The characteristics and section location of social science references

The last analyses look into the cited social science references and investigate 1) what kind of social science publications are cited by life science research, and 2) where and for what purpose they are cited.

3.3.1 The characteristics of social science references: interdisciplinarity and citation impact

The endeavor of life science publications citing social sciences can be regarded as interdisciplinary knowledge-seeking practices that are exploratory and unpredictable, compared to citing knowledge from their own domain. When crossing disciplinary borders, what kind of knowledge did life science research retrieve from social sciences?

We first investigated the level of interdisciplinarity of cited social science publications, quantified as the relative diversity of disciplinary categories in references. We found that the cited social science publications are on average associated with a higher level of interdisciplinarity within their disciplines (Figure 5, left). It means that the majority of social science citers are citing the more atypical and interdisciplinary social science knowledge. On the other hand, we also see that *PLoS Medicine* and *PLoS Pathogens* cited a noticeable proportion of social science knowledge that is rather typical and disciplinary. We briefly discuss the connotation of these two scenarios and illustrate them using examples from information science as a cited social science discipline. In the first scenario (the cited are more interdisciplinary), one possible explanation is that the cited social science research works on similar interdisciplinary topics that span both social sciences and life sciences, hence high interdisciplinarity. Therefore, life science research cites this category of social science papers to find interdisciplinary companions that share similar research interests and work on the same topic from a different angle, that is to say, transdisciplinary homogeneity. For instance, the information needs or information behavior of patients and clinical professionals are studied by researchers from both library and information science (Jerome et al., 2001) and clinical research (Crangle et al., 2018). The common interest encouraged the emergence of clinical informatics and dedicated research centers. Under this category, such citations should be categorized as interdisciplinary but two interdisciplinary actors cite each other reciprocally. The latter scenario (the cited are less interdisciplinary) may capture interdisciplinary activities that aim to retrieve theories or methods that emerge and develop in extramural fields. Their agenda may be to harvest what the cited fields are good at and what they don't currently possess, i.e. interdisciplinary heterogeneity (Henneke & Lüthje, 2007). For instance, publications on indexing (Deerwester et al., 1990) or term weighting (Salton & Buckley, 1988) were cited by biological research that studied biomedical relationships (Chadwick et al., 2012; Percha & Altman, 2015) and genetic similarity (Murray et al., 2017). Bibliometric tools (C. Chen, 2006) and techniques, such as text mining (Smalheiser, 2012), are also cited by natural science research to assist in mapping knowledge domains (Breugelmans et al., 2015; Gao et al., 2017) or knowledge discovery (Haagen et al., 2013).

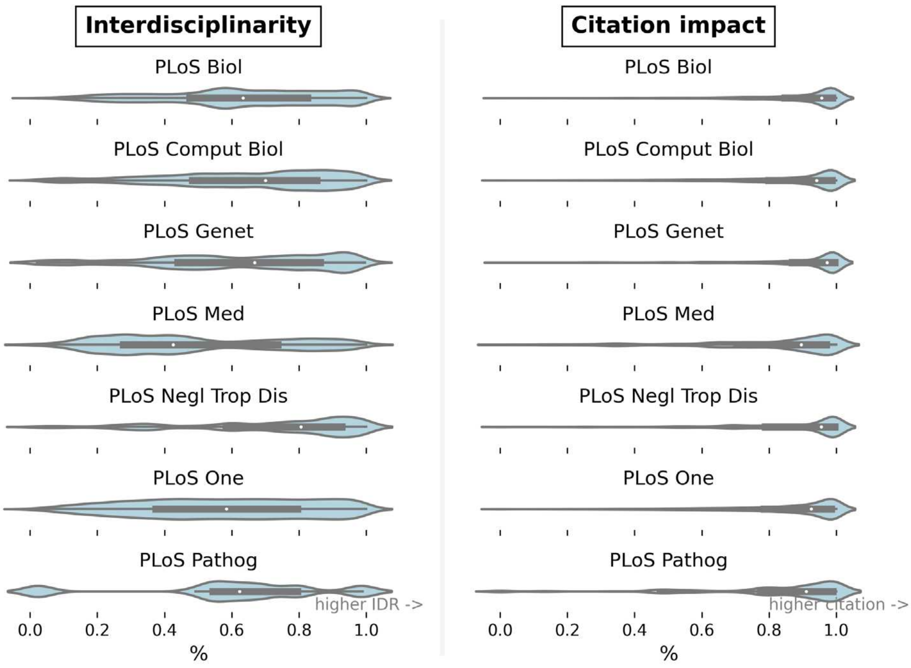

Figure 5. The interdisciplinarity and impact level of cited social science publications (normalized as percentile).

Regarding the citation impact, it is clear that the cited social science publications exhibit a very high level of citation impact compared to other publications from the same field and published year (Figure 5, right). It is possible that life science research intentionally selected research that had obtained high citations and regarded them as "good" research to ensure the quality of information seeking while lacking expertise in social sciences. Another hypothesis is that the cited social science research achieved a high impact since it is cited and adopted by multiple fields. Nonetheless, life science research is obtaining knowledge primarily from the most influential works in social sciences.

3.3.2 The section location of social science references

The section locations of references are indicative of the purpose of knowledge borrowing in the design of research by authors. Of all the references we recognized in our dataset, most references reside in the Introduction (37.5%), and Discussion (36.9%), respectively. This is consistent with findings from Bertin et al. (2016) that the Introduction and Discussion sections host the most references. 13.1% of references can be found in the Methodology section, whereas the Results section holds the least (12.4%). For the classified social science references (venue classification), we find them most frequently in the Introduction (41.5%) and Discussion (39.7%), and least frequently in the Results section (5.1%). 13.7% of references are cited in the Methodology section of life science publications. Considering social science disciplines excluding Psychology, we find 22.1% of references in the Methodology section. Over the observed periods, we found a greater proportion of social science references in the Introduction section, which is consistent with the pattern of all the references (Figure 6a). More than 80% of references are cited only in one section. 15.1% of social science references are cited in two sections (11.7% for social science disciplines except Psychology), which is similar to all other references (15.5%). Regardless of the discipline of references, the Introduction and Discussion section share the most references – all social science: 10.9%, social science except Psychology: 6.6%, non-social science: 8.8%.

Furthermore, we delve into the distribution of references in sections for different social science disciplines to uncover if they deliver different functions in their relationships with life sciences. In Figure 6b, we highlighted the percentages of references from various social science disciplines in two sections particularly, namely Introduction and Method. For most social science disciplines, they contributed a higher percentage of references in the Introduction section than the average level. Surprisingly, more than half of the social science disciplines contribute a higher percentage of references in the Methodology section than the average level. Two disciplines contribute more than 50% of references in Methodology, namely *Social Science Method*, and *Econometrics*. *Information & Library Science*, *Finance*, and *Sociology* also provided 24 to 34% of their references to the Methodology section in our life science publication set.

In summary, we observe that publications from social sciences not only support life science research in sections that primarily discuss backgrounds, theories, and implications, namely, Introduction and Discussion but also participate in the research design and methodology-related sections. Nonetheless, various social science disciplines play different roles in exporting knowledge to life sciences, some contribute primarily method-related knowledge yet some are more helpful in the theoretical aspects of the research.

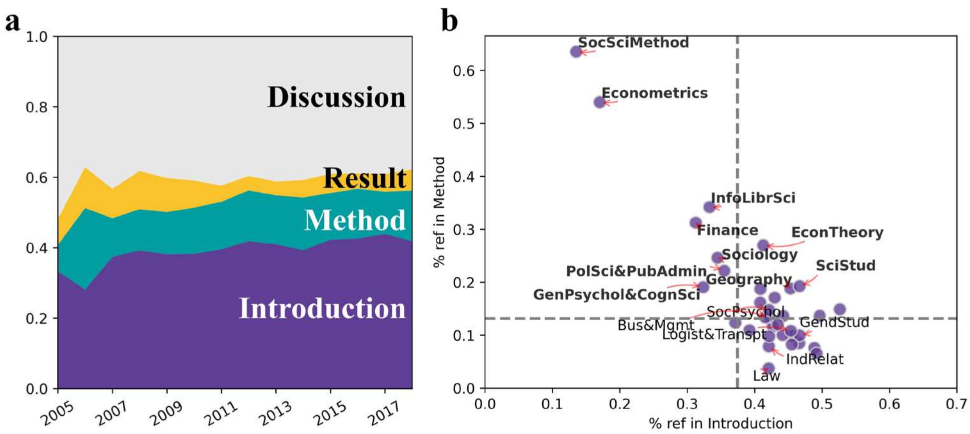

Figure 6. The section location of social science references. (a) temporal evolution in the proportion of social science references in different sections; (b) the proportion of references in the Introduction and Method for various social science disciplines. The vertical dashed line denotes the proportion in the Method section for all references (13.1%), while the horizontal line denotes that of the Introduction section (37.5%). Ten disciplines with the highest percentage of references in the Method section are highlighted in bold.

4 CONCLUSION

This study takes life science publications in seven journals in PLoS collections as an example and attempts to characterize the volume and trajectory of citations from life sciences to social sciences. Three groups of questions are explored: To what extent do life sciences cite social science? What kind of research in life sciences is citing social sciences and could one find any characteristics of these citers? What are the characteristics of cited social science knowledge, and for what purpose? To tackle these questions, we designed three groups of analyses and hope to delineate the characteristics of various actors in this set of knowledge diffusion and gain a better understanding of the role of social sciences with regard to life sciences.

Our results show that natural science publications in our dataset are not only citing social science knowledge but also directly citing their journals and researchers. From 2005 to 2018, a greater percentage of life science publications cited social sciences, which also accounts for more references. We estimate that social science disciplines influence around 15–19% of life science publications and accounted for 1.1–1.5% of references in 2018. If excluding Psychology, the rest of the social science disciplines influenced around 8–14% of life science publications and accounted for 0.4–0.6% of references in 2018. When pinpointing the cognitive location of these social science citers, we see that they appeared in the majority of topics (clusters) within life science journals and can be found in both the periphery and central topics in their respective disciplines. Some journals, for example, *PLoS Genetics* and *PLoS Biology*, exhibit a greater concentration of social science citers in a few topics, whereas that of other journals is more scattered. The cited social science publications are associated with the greatest citation impact within their coeval peers in their disciplines and seem to exhibit either the highest or the lowest level of interdisciplinarity. Zooming in on the reference section to infer citation purpose, we see that Introduction and Discussion hold the most social science references. Many social science disciplines are cited more heavily also in the Methodology sections than the average level: more than 50% of references from Social Science Method and Econometrics are found in the Methodology section, and 24–34% for Information & Library Science, Finance, and Sociology. We show that social sciences contribute to both the theoretical and methodological aspects of some life science research.

Through empirical evidence from publications, we discuss the diverse contributions from social sciences to life sciences and the growing affinity between the two branches. Although focusing on citations in one direction, we would argue the narrowing gap is achieved by both parties. On the one hand, life sciences are increasingly citing social sciences; on the other hand, social sciences are also becoming more interdisciplinary (Zhou et al., 2022) so that more research merges with life science and is more "citable" by the life science community. To further bridge the two branches, researchers from both sides should gain a better understanding of each other's research practices and create favourable conditions for communication and collaboration. We discuss one example. From our case studies and as pointed out by Porter (1994), statistics and data-oriented knowledge become one of the weightiest and most commutable factors that facilitate the knowledge diffusion between natural and social sciences. Data on the socioeconomic status of a population could be utilized in natural science research to factor human activities in natural systems. However, data sharing is reported to be the least practiced or preferred among social science researchers and the least involved by research institutions for a formally established data-managing process (Tenopir et al., 2011). In social scientists' defense, it could be that their data may be "sensitive to human subject of ethical guidelines" (Tenopir et al., 2011, p. 6), which impedes easy and unconditional sharing of data. Nonetheless, it is necessary to propose standardized and suitable data publishing/sharing protocols or tools for social sciences and techniques of merging natural and social data, for instance, through a uniform scaling system (Richter et al., 2022). In addition to advocating for communication between scientists, we encourage research institutions and science policymakers to think deeply about infrastructure building for better interdisciplinary knowledge diffusion.

Our study's approach to characterize the presence and influence of social science research could be applied, at least partially, by research management entities for diverse purposes. For instance, using similar analyses, universities may benchmark their research portfolio against our findings to gain insight in the extent to which their life science research incorporates social science knowledge. For instance, do STEM universities exhibit deficit in social perspectives in research given the lack of social science departments? If so, what initiatives can be taken to equip their researchers with the necessary social science knowledge to avoid being trapped in disciplinary silos or out-of-touch with potential social issues and solutions? Admittedly, implementing all the analyses discussed in this article within research management frameworks could pose challenges for smaller institutions, primarily due to constraints like data access and computational resources. However, recent advancements in open research data, spearheaded by initiatives such as OpenAlex (Priem et al., 2022), could significantly facilitate this process. In addition, the life science research community in general, and its various communication actors, e.g., journals, may also want to survey and monitor emerging social science knowledge that has been employed or is potentially useful to their research. Such an initiative may speed up and broaden the spread of up-to-date, high-quality social science knowledge, enabling researchers from all status to seize interdisciplinary opportunities. As the emphasis on the social responsibilities of research and engagement with major societal challenges grows, it's crucial for natural science researchers to have a deeper understanding of their ultimate end-users: society at large, and the scientific principles that govern it.

Finally, we must stress that this study is limited to the dataset we adopted, which covers mostly the animated branches of natural sciences, namely, biomedical and life sciences. These disciplines may exhibit greater affinity with social sciences in nature since they research human-related topics. It is possible that disciplines less covered in this study, such as Physics and Chemistry, cite fewer references from social sciences. Results from this study may represent the upper limit of natural sciences citing social sciences. In future research, we may expand our analyses to other branches of natural science and depict more interesting aspects of interdisciplinary flow, such as the qualitative/quantitative orientations of the citing and the cited publications, and the exact trajectory and actor of knowledge diffusion.

5 ACKNOWLEDGEMENTS

Hongyu Zhou, Raf Guns, and Tim C.E. Engels thank the Flemish Government for its support to the Centre for R&D Monitoring (ECOOM), which made possible this research as part of the program on the study of interdisciplinarity and impact. The opinions in the paper are the authors' and not necessarily those of the government. Lin Zhang acknowledges the support of the National Natural Science Foundation of China (grant no. 72374160).

6 REFERENCES

Anderson, D. R., Bryant, J., Murray, J. P., Rich, M., Rivkin, M. J., & Zillmann, D. (2006). Brain Imaging-An Introduction to a New Approach to Studying Media Processes and Effects. *Media Psychology*, *8*(1), 1–6. https://doi.org/10.1207/S1532785XMEP0801_1

Archambault, É., Beauchesne, O. H., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. *Proceedings of the 13th International Conference of the International Society for Scientometrics and Informetrics*, *13*, 66–77.

Bennett, N. J., Roth, R., Klain, S. C., Chan, K. M. A., Clark, D. A., Cullman, G., Epstein, G., Nelson, M. P., Stedman, R., Teel, T. L., Thomas, R. E. W., Wyborn, C., Curran, D., Greenberg, A., Sandlos, J., & Veríssimo, D. (2017). Mainstreaming the social sciences in conservation. *Conservation Biology*, *31*(1), 56–66. https://doi.org/10.1111/cobi.12788

Bertin, M., & Atanassova, I. (2014). A Study of Lexical Distribution in Citation Contexts through the IMRaD Standard. *BIR@ECIR*.

Bertin, M., Atanassova, I., Gingras, Y., & Larivière, V. (2016). The invariant distribution of references in scientific articles. *Journal of the Association for Information Science and Technology*, *67*(1), 164–177. https://doi.org/10.1002/asi.23367

Bonaccorsi, A. (2022). An epistemic approach to research assessment in the social sciences. In *Handbook on Research Assessment in the Social Sciences* (pp. 14–47). Edward Elgar Publishing. https://www.elgaronline.com/view/edcoll/9781800372542/9781800372542.00009.xml

Breugelmans, J. G., Makanga, M. M., Cardoso, A. L. V., Mathewson, S. B., Sheridan-Jones, B. R., Gurney, K. A., & Mgone, C. S. (2015). Bibliometric assessment of European and sub-Saharan African research output on poverty-related and neglected infectious diseases from 2003 to 2011. *PLoS Neglected Tropical Diseases*, *9*(8), e0003997.

Calvert, J., & Martin, P. (2009). The role of social scientists in synthetic biology. *EMBO Reports*, *10*(3), 201–204.

Camerer, C. F., Loewenstein, G., & Prelec, D. (2004). Neuroeconomics: Why Economics Needs Brains. *The Scandinavian Journal of Economics*, *106*(3), 555–579. https://doi.org/10.1111/j.0347-0520.2004.00377.x

Campbell, S. J., Nery, S. V., Wardell, R., D'Este, C. A., Gray, D. J., McCarthy, J. S., Traub, R. J., Andrews, R. M., Llewellyn, S., Vallely, A. J., Williams, G. M., & Clements, A. C. A. (2017). Water, Sanitation and Hygiene (WASH) and environmental risk factors for soil-transmitted helminth intensity of infection in Timor-Leste, using real time PCR. *PLOS Neglected Tropical Diseases*, *11*(3), e0005393. https://doi.org/10.1371/journal.pntd.0005393

Chadwick, W., Martin, B., Chapter, M. C., Park, S.-S., Wang, L., Daimon, C. M., Brenneman, R., & Maudsley, S. (2012). GIT2 Acts as a Potential Keystone Protein in Functional Hypothalamic Networks Associated with Age-Related Phenotypic Changes in Rats. *PLOS ONE*, *7*(5), e36975. https://doi.org/10.1371/journal.pone.0036975

Chen, C. (2006). CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for Information Science and Technology*, *57*(3), 359–377.

Chen, X., Mohr, K., & Galea, J. M. (2017). Predicting explorative motor learning using decision-making and motor noise. *PLOS Computational Biology*, *13*(4), e1005503. https://doi.org/10.1371/journal.pcbi.1005503

Collins, R. (1994). Why the social sciences won't become high-consensus, rapid-discovery science. *Sociological Forum*, *9*(2), 155–177. https://doi.org/10.1007/BF01476360

Crangle, C. E., Bradley, C., Carlin, P. F., Esterhay, R. J., Harper, R., Kearney, P. M., McCarthy, V. J. C., McTear, M. F., Savage, E., Tuttle, M. S., & Wallace, J. G. (2018). Exploring patient information needs in type 2 diabetes: A cross sectional study of questions. *PLOS ONE*, *13*(11), e0203429. https://doi.org/10.1371/journal.pone.0203429

Deerwester, S., Dumais, S. T., Furnas, G. W., Landauer, T. K., & Harshman, R. (1990). Indexing by latent semantic analysis. *Journal of the American Society for Information Science*, *41*(6), 391–407. https://doi.org/10.1002/(SICI)1097-4571(199009)41:6<391::AID-ASI1>3.0.CO;2-9

Filmer, D., & Pritchett, L. H. (2001). Estimating wealth effects without expenditure data—or tears: An application to educational enrollments in states of India. *Demography*, *38*(1), 115–132. https://doi.org/10.1353/dem.2001.0003

Gao, Y., Wang, Y., Zhai, X., He, Y., Chen, R., Zhou, J., Li, M., & Wang, Q. (2017). Publication trends of research on diabetes mellitus and T cells (1997–2016): A 20-year bibliometric study. *PLOS ONE*, *12*(9), e0184869. https://doi.org/10.1371/journal.pone.0184869

Gernsbacher, M. A., & Kaschak, M. P. (2003). Neuroimaging Studies of Language Production and Comprehension. *Annual Review of Psychology*, *54*(1), 91–114. https://doi.org/10.1146/annurev.psych.54.101601.145128

Glänzel, W., & Schoepflin, U. (1999). A bibliometric study of reference literature in the sciences and social sciences1Parts of this study have been presented at the 5th International Conference on Scientometrics and Informetrics, held in River Forest, Illinois, June 7–10, 1995.1. *Information Processing & Management*, *35*(1), 31–44. https://doi.org/10.1016/S0306-4573(98)00028-4

Grover, A., & Leskovec, J. (2016). node2vec: Scalable feature learning for networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 855–864.

Gwizdka, J., Mostafa, J., Moshfeghi, Y., Bergman, O., & Pollick, F. E. (2013). Applications of neuroimaging in information science: Challenges and opportunities. *Proceedings of the American Society for Information Science and Technology*, *50*(1), 1–4. https://doi.org/10.1002/meet.14505001016

Haagen, H. H. H. B. M. van, Hoen, P. A. C. 't, Mons, B., & Schultes, E. A. (2013). Generic Information Can Retrieve Known Biological Associations: Implications for Biomedical Knowledge Discovery. *PLOS ONE*, *8*(11), e78665. https://doi.org/10.1371/journal.pone.0078665

Heise, D. R., & Simmons, R. G. (1985). Some Computer-Based Developments in Sociology. *Science*, *228*(4698), 428–433.

Henneke, D., & Lüthje, C. (2007). Interdisciplinary Heterogeneity as a Catalyst for Product Innovativeness of Entrepreneurial Teams. *Creativity and Innovation Management*, *16*(2), 121–132. https://doi.org/10.1111/j.1467-8691.2007.00426.x

Hill, M. O. (1973). Diversity and evenness: A unifying notation and its consequences. *Ecology*, *54*(2), 427–432.

Huang, M., & Chang, Y. (2008). Characteristics of research output in social sciences and humanities: From a research evaluation perspective. *Journal of the American Society for Information Science and Technology*, *59*(11), 1819–1828. https://doi.org/10.1002/asi.20885

Jaffe, K. (2014). Social and Natural Sciences Differ in Their Research Strategies, Adapted to Work for Different Knowledge Landscapes. *PLOS ONE*, *9*(11), e113901. https://doi.org/10.1371/journal.pone.0113901

Jerome, R. N., Giuse, N. B., Wilder Gish, K., Sathe, N. A., & Dietrich, M. S. (2001). Information needs of clinical teams: Analysis of questions received by the Clinical Informatics Consult Service. *Bulletin of the Medical Library Association*, *89*(2), 177–185.

Kagan, J. (2009). *The Three Cultures: Natural Sciences, Social Sciences, and the Humanities in the 21st Century*. Cambridge University Press.

Kai-Ineman, D., & Tversky, A. (1979). Prospect theory: An analysis of decision under risk. *Econometrica*, *47*(2), 363–391.

Knopp, S., Mohammed, K. A., Stothard, J. R., Khamis, I. S., Rollinson, D., Marti, H., & Utzinger, J. (2010). Patterns and Risk Factors of Helminthiasis and Anemia in a Rural and a Peri-urban Community in Zanzibar, in the Context of Helminth Control Programs. *PLOS Neglected Tropical Diseases*, *4*(5), e681. https://doi.org/10.1371/journal.pntd.0000681

Kuhn, T. S. (2012). *The structure of scientific revolutions*. University of Chicago press.

Kulczycki, E., Guns, R., Pölönen, J., Engels, T. C. E., Rozkosz, E. A., Zuccala, A. A., Bruun, K., Eskola, O., Starčič, A. I., Petr, M., & Sivertsen, G. (2020). Multilingual publishing in the social sciences and humanities: A seven-country European study. *Journal of the Association for Information Science and Technology*, *71*(11), 1371–1385. https://doi.org/10.1002/asi.24336

Larivière, V., Archambault, É., Gingras, Y., & Vignola-Gagné, É. (2006). The place of serials in referencing practices: Comparing natural sciences and engineering with social sciences and humanities. *Journal of the American Society for Information Science and Technology*, *57*(8), 997–1004. https://doi.org/10.1002/asi.20349

Larivière, V., Gingras, Y., & Archambault, É. (2006). Canadian collaboration networks: A comparative analysis of the natural sciences, social sciences and the humanities. *Scientometrics*, *68*(3), 519–533. https://doi.org/10.1007/s11192-006-0127-8

Medical and Social. (1958). Medicine and Social Science. Editorial. *New England Journal of Medicine*, *258*(8), 400–401. https://doi.org/10.1056/NEJM195802202580813

Meloni, M. (2014). How Biology Became Social, and What it Means for Social Theory. *The Sociological Review*, *62*(3), 593–614. https://doi.org/10.1111/1467-954X.12151

Moon, K., & Blackman, D. (2014). A guide to understanding social science research for natural scientists. *Conservation Biology: The Journal of the Society for Conservation Biology*, *28*(5), 1167–1177. https://doi.org/10.1111/cobi.12326

Murray, K. D., Webers, C., Ong, C. S., Borevitz, J., & Warthmann, N. (2017). kWIP: The k-mer weighted inner product, a de novo estimator of genetic similarity. *PLOS Computational Biology*, *13*(9), e1005727. https://doi.org/10.1371/journal.pcbi.1005727

Neumann, Y. (1977). Standards of Research Publication: Differences between the Physical Sciences and the Social Sciences. *Research in Higher Education*, *7*(4), 355–367.

Percha, B., & Altman, R. B. (2015). Learning the Structure of Biomedical Relationships from Unstructured Text. *PLOS Computational Biology*, *11*(7), e1004216. https://doi.org/10.1371/journal.pcbi.1004216

Pickersgill, M. (2013). The social life of the brain: Neuroscience in society. *Current Sociology. La Sociologie Contemporaine*, *61*(3), 322–340. https://doi.org/10.1177/0011392113476464

Porter, T. M. (1994). From Quetelet to Maxwell: Social Statistics and the Origins of Statistical Physics. In I. B. Cohen (Ed.), *The Natural Sciences and the Social Sciences: Some Critical and Historical Perspectives* (pp. 345–362). Springer Netherlands. https://doi.org/10.1007/978-94-017-3391-5_11

Price, D. J. (1970). Citation measures of hard science, soft science, technology, and nonscience. *Communication among Scientists and Engineers*, *1*, 3r22.

Priem, J., Piwowar, H., & Orr, R. (2022, September 7). OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *In N. Robinson-Garcia, D. Torres-Salinas, & W. Arroyo-Machado (Eds.)*. 26th International Conference on Science and Technology Indicators, STI 2022 (sti22193), Granada, Spain. https://doi.org/10.5281/zenodo.6936227

Richter, I., Roberts, B. R., Sailley, S. F., Sullivan, E., Cheung, V. V., Eales, J., Fortnam, M., Jontila, J. B., Maharja, C., Nguyen, T. Ha., Pahl, S., Praptiwi, R. A., Sugardjito, J., Sumeldan, J. D. C., Syazwan, W. M., Then, A. Y., & Austen, M. C. (2022). Building bridges between natural and social science disciplines: A standardized methodology to combine data on ecosystem quality trends. *Philosophical Transactions of the Royal Society B: Biological Sciences*, *377*(1854), 20210487. https://doi.org/10.1098/rstb.2021.0487

Salton, G., & Buckley, C. (1988). Term-weighting approaches in automatic text retrieval. *Information Processing & Management*, *24*(5), 513–523. https://doi.org/10.1016/0306-4573(88)90021-0

Shen, Z., Chen, F., Yang, L., & Wu, J. (2019). Node2vec Representation for Clustering Journals and as A Possible Measure of Diversity. *Journal of Data and Information Science*, *4*(2), 79–92. https://doi.org/10.2478/jdis-2019-0010

Smalheiser, N. R. (2012). Literature-based discovery: Beyond the ABCs. *Journal of the American Society for Information Science and Technology*, *63*(2), 218–224. https://doi.org/10.1002/asi.21599

Stefaniak, B. (2001). International Co-operation in Science and in Social Sciences as Reflected in Multinational Papers Indexed in SCI and SSCI. *Scientometrics*, *52*(2), 193–210. https://doi.org/10.1023/A:1017959605687

Stephan, K. E., Kasper, L., Harrison, L. M., Daunizeau, J., den Ouden, H. E. M., Breakspear, M., & Friston, K. J. (2008). Nonlinear dynamic causal models for fMRI. *NeuroImage*, *42*(2), 649–662. https://doi.org/10.1016/j.neuroimage.2008.04.262

Tahamtan, I., & Bornmann, L. (2019). What do citation counts measure? An updated review of studies on citations in scientific documents published between 2006 and 2018. *Scientometrics*, *121*(3), 1635–1684. https://doi.org/10.1007/s11192-019-03243-4

Tenopir, C., Allard, S., Douglass, K., Aydinoglu, A. U., Wu, L., Read, E., Manoff, M., & Frame, M. (2011). Data Sharing by Scientists: Practices and Perceptions. *PLOS ONE*, *6*(6), e21101. https://doi.org/10.1371/journal.pone.0021101

Thijs, B. (2020). Using neural-network based paragraph embeddings for the calculation of within and between document similarities. *Scientometrics*, *125*(2), 835–849. https://doi.org/10.1007/s11192-020-03583-6

Van der Maaten, L., & Hinton, G. (2008). Visualizing data using t-SNE. *Journal of Machine Learning Research*, *9*(11).

Wang, K., Shen, Z., Huang, C., Wu, C.-H., Dong, Y., & Kanakia, A. (2020). Microsoft Academic Graph: When experts are not enough. *Quantitative Science Studies*, *1*(1), 396–413. https://doi.org/10.1162/qss_a_00021

Wen, X., Yao, L., Liu, Y., & Ding, M. (2012). Causal Interactions in Attention Networks Predict Behavioral Performance. *Journal of Neuroscience*, *32*(4), 1284–1292. https://doi.org/10.1523/JNEUROSCI.2817-11.2012

Xu, X., Tan, A. M., & Zhao, S. X. (2015). Funding ratios in social science: The perspective of countries/territories level and comparison with natural sciences. *Scientometrics*, *104*(3), 673–684. https://doi.org/10.1007/s11192-015-1633-3

Zhang, L., Sun, B., Chinchilla-Rodríguez, Z., Chen, L., & Huang, Y. (2018). Interdisciplinarity and collaboration: On the relationship between disciplinary diversity in departmental affiliations and reference lists. *Scientometrics*, *117*(1), 271–291. https://doi.org/10.1007/s11192-018-2853-0

Zhou, H., Guns, R., & Engels, T. C. E. (2022). Are social sciences becoming more interdisciplinary? Evidence from publications 1960–2014. *Journal of the Association for Information Science and Technology*, *73*(9), 1201–1221. https://doi.org/10.1002/asi.24627

Zhou, H., Guns, R., & Engels, T. C. E. (2023). Towards indicating interdisciplinarity: Characterizing interdisciplinary knowledge flow. *Journal of the Association for Information Science and Technology*, *74*(11), 1325–1340. https://doi.org/10.1002/asi.24829

APPENDIX

- **A1. Mapping section names to unified section structures (IMRaD)**

To be able to pinpoint the location of social science references for all articles, one must have a unified section structure such as IMRaD (Introduction, Methodology, Results and Discussion) to make comparisons possible. We therefore mapped section names into six unified sections, namely, **I**ntroduction, **M**ethodology, **R**esults, **D**iscussion, SI, and **O**thers. The mapping is conducted in three steps.

Firstly, for each unprocessed section name, we searched for keywords listed in Table A1 and mapped them to according unified sections. Secondly, if one unprocessed section name is mapped with multiple unified sections, we perform the following two tasks. Priorities are given to 'SI' and Results since the search keywords are more specific for them. Next, we compared the currently mapped section names with the last adjacent section and removed what had already occurred in the previous section. Thirdly, the unmatched cases are handled. If it occurs at the beginning of the document, we classify it as Introduction; otherwise, it is classified as the same mapped section as its previous section.

Table A1. Keywords for mapping section names to unified section structures

Mapped section	**Keywords**
INTRODUCTION	introduction, background, motivation
METHODOLOGY	method(s), material(s), methodology, dataset, database
RESULTS	result(s)
DISCUSSION	discussion(s), conclusion(s), limitation(s), summary
SI	supporting, supplementary, appendix, competing interests, data availability, corresponding author
OTHERS	correction(s)

- **A2. Extracting departmental affiliations of authors and mapping to disciplines**

To recognize researchers who work in departments from social sciences, for each reference in our dataset, we obtained the address line of the first author and processed the texts in the following steps.

First, the texts are segmented by punctuation marks such as comma (,), semicolon (;), and forward slash (/);

Second, we identify secondary institutions, i.e., departmental affiliation, by searching a list of cue words in segmented texts. The following words are used to represent secondary institutions: *department, school, center/centre, college, institution/institute, lab, association, faculty, division, hospital, unit, program, discipline, clinic, group, and team*. For texts that contain "*college*" and "*institute*", we only consider it as the secondary institution if it occurs at the beginning of the address line so that institutions such as "*College of William & Mary*" or "*Massachusetts Institute of Technology*" wouldn't be recognized as secondary institutions.

Finally, we used an improved thesaurus from our previous studies (Zhang et al., 2018) and classified secondary institutions into 13 disciplines: *Medicine, Biology, Agriculture & Environment, Social Science, Engineering, Chemistry, Geosciences & Space Sciences, Economy & Management, Physics, Mathematics, Computer Science & Information Technology, Psychology, Arts & Humanities.*

We used three discipline labels, namely *Social Science*, *Economy & Management*, and *Psychology*, to denote social sciences in the first author's affiliation. We classify a secondary institution as social sciences only if it contains cue words from the above three discipline labels in our thesaurus. Therefore, secondary institutions with interdisciplinary traces, for example, the Center for Clinical and Genetic Economics, will not be classified as social sciences in this study.

The above processing procedures were conducted using the VantagePoint software (www.thevantagepoint.com).

- **A3. List of life science disciplines (MAG L1 classification)**

Biological and Environmental Sciences (22 fields):

Agricultural Science
Agronomy
Animal Science
Biochemistry
Bioinformatics
Biology
Botany
Cell Biology
Ecology
Environmental Health
Evolutionary Biology
Food Science
Genetics
Horticulture
Immunology
Microbiology
Molecular Biology
Neuroscience
Oceanography
Paleontology
Physiology
Zoology

Medical and Health Sciences (43 fields):

Alternative Medicine
Anatomy
Andrology
Anesthesia
Biomedical Engineering
Cancer Research
Cardiology
Clinical Psychology
Computational Biology
Dentistry
Emergency Medicine
Endocrinology
Family Medicine
Gastroenterology
General Surgery
Genetics
Gynecology
Intensive Care Medicine
Internal Medicine
Medical Emergency
Medical Physics
Medicinal Chemistry
Nursing
Nuclear Medicine
Obstetrics
Oncology
Ophthalmology
Optometry
Pathology
Pediatrics
Pharmacology
Physical Medicine & Rehabilitation
Physical Therapy
Psychiatry
Psychoanalysis
Psychotherapy
Radiology
Surgery
Toxicology
Traditional Medicine
Urology
Veterinary Medicine
Virology

Agricultural, Earth, and Marine Sciences (10 fields):

Agricultural Science
Agronomy
Animal Science
Botany
Ecology
Environmental Health
Food Science
Horticulture
Oceanography
Paleontoloy