Science maps created exclusively with section-specific citations: Which topics emerge?

Juan Pablo Bascur[1,2], Suzan Verberne[2]

[1]Centre for Science and Technology Studies, Leiden University, The Netherlands
[2]Leiden Institute of Advanced Computer Science, Leiden University, The Netherlands

Abstract

Traditional science maps represent research topics by clustering documents, yet this process systematically favors some kinds of topics over others. If such topical emphasis could be influenced in a controlled way, science maps could be adapted to better support different use cases. In this paper, we investigate whether section-specific citation networks, derived from the Introduction, Methods, Results, or Discussion sections, can be used to improve the clustering effectiveness of certain topics relative to others. We evaluate differences in clustering effectiveness between section-specific networks and a section-agnostic baseline network for topics belonging to 122 MeSH-based topic categories. Our largest network contains 7,507,751 documents and 36,286,575 citations. The results show that section-specific networks exhibit lower clustering effectiveness than the generic network for almost all topic categories, and that this reduced performance is likely related to the smaller number of citations rather than to intrinsic properties of section-specific citations. At the same time, we observe that different sections emphasize different kinds of topics, like social sciences in the Introduction, anatomy and medical techniques in the Discussion, and chemical substances in the Results. These findings indicate that while section-specific citation networks can shift topical salience in science maps, they are more effective as a complementary signal than as a standalone basis for map construction.

1. Introduction

Science maps are visualization techniques intended to represent the thematic organization of a collection of academic documents. They are typically used for literature analysis [1], field delimitation, research policy, and enhanced document browsing [2]. Typical practice for creating science maps is first to create a network of academic documents where the links reflect shared characteristics such as bibliographic metadata, then to cluster together the documents that are well connected, and finally to summarize the contents of these clusters.

In an ideal science map, each cluster would represent a single academic topic, and every document from each topic would be assigned to its corresponding cluster. In clustering evaluation, the extent to which a clustering solution (i.e. the set of clusters) approaches an underlying ground truth is called clustering effectiveness [3]. A perfect clustering effectiveness can never be achieved in real science maps because documents have multiple topics. In our previous work [4], we evaluated citation-based and text-similarity-based science maps for the biomedical domain, and found that there is a systematic difference between the clustering effectiveness of topics depending on the category of topic. For example, topics related to organisms or diseases have the highest effectiveness, while topics related to geographical locations have the lowest. In a later work [5], we found that basing the science maps on other kinds of attributes, beyond citation and text similarity, increases or decreases the clustering effectiveness of different kinds of topics. For example, we found that a science map based on co-authorship has much higher clustering effectiveness for topics related to geographical locations, relative to one based on text similarity, while at the same time also having a lower clustering effectiveness for the other kinds of topics that we evaluated.

In the current paper, we continue this line of research, but this time with maps created exclusively with citations that come from a given section of the academic document, be it Introduction, Methods,

SCOLIA '26: Second International Workshop on Scholarly Information Access (SCOLIA), April 2, 2026, Delft, The Netherlands
juanpablobascurcifuentes@gmail.com (J. P. Bascur)
0000-0002-4077-1024 (J. P. Bascur); 0000-0002-9609-9505 (S. Verberne)

results, or Discussion. This kind of analysis has become recently possible at a large scale due to Clarivate systematically providing this information in the Web of Science [6]. We specifically address the following research question: *For which topic categories does the clustering effectiveness increase or decrease when we create a science map using only the citations that come from a given section, compared with a science map that uses all the citations*?

Our contributions are the following:

- We present the first large-scale, systematic evaluation of science maps constructed exclusively from citations originating in a single document section (Introduction, Methods, Results, or Discussion), and compare their clustering effectiveness with that of section-agnostic citation maps across a wide range of topics.
- We conduct this evaluation across 122 MeSH-based topic categories and multiple network constructions with up to 7 million documents and 36 million citations, evaluated across 140 clustering resolutions and three coverage levels, allowing us to draw robust, category-level conclusions that are representative of multiple clustering granularities.
- We show that, despite their lower overall effectiveness, different section-specific citation networks emphasize different kinds of topics, suggesting that section-specific citations can function as a complementary signal for influencing topic emergence in science maps.

2. Related work

The scientific field of *citation context analysis* studies how and why citations are used within academic documents, including how citation behavior differs across sections. Prior work in this field has shown that citation usage varies systematically between sections, suggesting that citations appearing in different parts of a document serve different roles. This motivates our expectation that science maps constructed from citations restricted to specific sections may improve the clustering effectiveness of particular kinds of topics. At the same time, it is important to note that citation behavior is not stable over time [7, 8] and that substantial differences exist across disciplines [9, 8].

Prior literature provides evidence for the following section-specific citation characteristics:

- **Introduction section**: Citations in the Introduction account for the largest share of in-text citations within a document [10] and disproportionately reference older, highly cited, and seminal works [11]. These citations are commonly used to position a study within an established body of literature.
- **Methods section**: Citations in the Methods section are, on average, older [12] and highly concentrated on a relatively small set of widely reused references [13]. Such citations tend to recur across documents whenever a given method, protocol, or tool is employed.
- **Results section**: Citations in the Results section are often used for direct empirical comparison with closely related studies.
- **Discussion section**: This section generally contains the fewest [10] and youngest [8] citations. These citations are frequently used to interpret findings and to relate the current work to broader themes or bodies of literature.

Several studies have also examined the effects of counting citations differently depending on the section in which they appear. On one hand, Thelwall [9] and Jeon and Jung [14] show that section-wise citation counting can lead to substantially different ranking outcomes. On the other hand, Donner et al. [15] find that such section-based citation counts, regardless of which section is considered, are not correlated with peer assessments of the cited document's quality. Taken together, these results indicate that while section-wise citation counting produces substantial differences in evaluation outcomes, these differences are not aligned with perceived academic quality.

Section-wise citation information has also been incorporated into machine learning models. It has also been used in citation recommendation systems, where recommendations are conditioned on the

Table 1
Initial data selection statistics

	Raw citations	Known section citations
Documents	34,587,864	9,911,919
Core documents	4,094,364	3,434,877
Citations	258,238,830	39,145,982
Citations between core documents	18,579,805	479,905
Citations with multiple sections	–	5,938,464
Citations with multiple sections between core documents	–	90,006

section in which a citation is to be placed [16, 17]. Liang et al. [18] train a language model for science mapping using citations weighted according to their section of origin. Their results show that excluding citations from individual sections leads to only small changes in overall map quality, although the highest performance is achieved by excluding citations from the Methods section. Their work is the most closely related to our own; however, the underlying goal differs. Whereas Liang et al. aim to optimize overall mapping quality through section weighting, our goal is to understand how restricting citations to a single section affects the clustering effectiveness of different kinds of topics, under the assumption that topics may be structurally incompatible.

3. Methods

3.1. Source of the networks

Our citation networks are built as extended direct citation networks [19]. These are direct citation networks with two types of documents: core documents, which are included in the science map, and non-core documents, which are excluded from the science map but are used during clustering. This approach allows the clustering to take into account that two documents can be connected through a third document, even if that third document is not part of the science map itself. This is particularly relevant for citation networks that are subsets of the global citation network (i.e., all documents within a given time period), which is the case for the networks we construct here.

Extended citation networks are created as follows. The network is undirected. For each core document, we add all non-core documents that are connected to at least two core documents. Edge weights are set to 1, node weights are set to 1 for core documents and 0 for non-core documents. The purpose of the latter is so that in the clustering algorithm we use, the Leiden algorithm, the quality function of a cluster is not reduced by adding non-core documents to that cluster.

All documents and citations that we use in the networks originate from the same initial set of documents and citations, defined as follows. We use the Web of Science (WoS) database from CWTS (version 2025.39) and select all documents published in the years 2016, 2017, 2018, and 2019. This publication window is chosen to allow documents to accumulate citations from documents published after 2019. From this set, we further select only documents that have a PubMed ID, using the DOI as a joining key and the PubMed database from CWTS (version 2025). This restriction is applied because, at a later stage, we map properties of MeSH terms from PubMed onto the documents for topic evaluation. Some WoS documents are associated with more than one PubMed ID; such documents are included only once. We refer to this collection as the initial set of core documents.

Using this initial set, we retrieve all WoS citations for which an initial core document appears either as the citing or the cited document, and for which the section of the citation is known. These are the only citations that we will use to construct the networks. If a document pair has citations in multiple sections, we keep all section labels, and during network creation we treat these citations as a single edge. Summary statistics of the initial data selection are reported in Table 1. WoS distinguishes four categories for the section of origin of citations: Introduction, Methods, Results, and Discussion. WoS does not publish the methodology used to disambiguate sections; however, in citation context analysis it is common practice to assign Background to Introduction and Conclusion to Discussion [20].

Table 2
Size of network sections

	Introduction	Methods	Results	Discussion
Core documents	2,950,267	1,232,253	1,398,134	2,605,770
Pure network documents	6,671,770	2,289,769	2,373,223	4,962,544
Baseline network documents	7,507,751	5,030,793	5,201,633	6,873,893
Pure to Baseline documents ratio	0.89	0.46	0.46	0.72
Pure network citations	20,063,427	3,789,303	4,337,552	13,193,498
Baseline network citations	36,286,575	23,180,813	25,372,983	33,724,671
Pure to Baseline citations ratio	0.55	0.16	0.17	0.39

3.2. Construction of the networks

For each section, we construct two networks: one containing only citations originating from that section, which we call the *Pure* network, and one containing citations from any section, which we call the *Baseline* network. The Baseline network serves as a reference to assess how using section-specific citations differs from using all citations. The sizes of all networks are reported in Table 2.

The construction of a Pure network proceeds as follows. We identify all initial citations belonging to the given section. From this set, we remove the citations with non-core documents that connect to a single core document. The remaining citations define the network, and the remaining core documents constitute the section-specific core document set.

To construct the Baseline network, we identify all initial citations that contain a section-specific core document. From this set, we remove the citations with non-section-specific-core documents that connect to a single section-specific core document. The remaining citations define the network, and only section-specific core documents are considered core documents. This means that some initial core documents can become non-core documents in this network. We do this so that the Pure and Baseline networks have exactly the same core documents, so we can later compare their clustering effectiveness.

3.3. Topics and topic categories

Topics are concepts that describe the content of a document, and a document can have any number of topics. Topic categories are a way of grouping topics together. For example, topics *Dog* and *Cat* belong to the topic category *Organisms*. We use the same method as in our previous work [5] to define which documents belong to which topics and to which topic categories these topics belong. In the current subsection, we will make a high level summary of this method and go into detail on how the current paper differs from this method.

The MeSH terms, which is a controlled vocabulary thesaurus used for indexing PubMed, are used to assign documents to topics and topics to topic categories. The MeSH tree is a structure that organizes the MeSH terms into a hierarchical structure, where tree nodes are subtopics of their parent node.

The topics in our work are the tree nodes of the MeSH tree. Assigning topics to a document works the following way: The document already has MeSH terms assigned (obtained from the in-house CWTS database of PubMed and MeSH version 2025). The MeSH terms are mapped to the MeSH tree, where they are transformed into their tree node version (there can be more than one tree node per MeSH term). Then, all the tree nodes that are ancestors of these tree nodes are also assigned to the document. We refer to a topic or a topic category (see below) using its tree node identifier followed by its MeSH term name (e.g., *[A01.456] Head*) so to differentiate between nodes of the same MeSH term.

The topic categories are the branch and first level of the MeSH tree (i.e., the two highest levels). All topics descendants of a given branch or first-level node are assigned to the corresponding topic category. We exclude the following topic categories:

- The entire branch *Publication Characteristics [V]*, including its first level, because it contains no documents.
- The categories *Humanities [K01]*, *Information Science [L01]*, *Persons [M01]*, and *Geographic Locations [Z01]*, because each of these is the only first-level MeSH term in its branch, making it operationally equivalent to the branch itself as a topic category.

In total, we obtained 122 topic categories.

We filter out some topics and topic categories to improve our evaluation, and we do independent filtering per section because each section has different core documents. First, we remove the topics that have a size (number of documents) outside a given range (see below for details), with the purpose of saving computational resources. Then we filter by redundancy, that is when topics that share a substantial number of documents (Jaccard similarity >= 0.5). From a set of redundant topics, we keep the one with the fewest documents and discard the rest. The final filter is topic category size, where we remove topic categories with less than 10 topics.

The size filtering range always has a minimum of 40 documents so to have meaningful topics. However, if after the redundancy filtering we have less than 50 topics, then we increase the maximum size and do the redundancy filtering again. We use the following maximum sizes: 250, 500, 100, 2000, 4000, 8000.

3.4. Clustering

We use the Leiden algorithm [21] for clustering, which is commonly applied in the construction of science maps. The algorithm requires the user to specify a resolution parameter, which controls cluster granularity, with higher resolution values leading to smaller clusters. For each network, we generate 140 clustering solutions using 140 different resolution values. The highest resolution is set to 10, and each subsequent resolution value is obtained by multiplying the previous one by a factor of $7/8$ (e.g., 10, 8.75, 7.65625, $\ldots$).

This geometric progression was chosen because it provides a smooth transition from clustering solutions in which most nodes are isolated to solutions in which most nodes belong to a single cluster, while keeping the total number of resolutions sufficiently low to allow computation within a reasonable time frame. In practice, clustering all networks for a single section takes approximately one day.

All clustering is performed using the Python implementation of the Leiden algorithm provided by the `igraph` library [22]. All clustering solutions generated across the different resolution values are retained and used in the subsequent evaluation and comparison analyses.

3.5. Evaluation

Our evaluation framework follows the methodology introduced in our previous work [5]. In this section we make a high level description of how we evaluate clustering effectiveness, and the formal definitions and implementation details are provided in [5].

The clustering effectiveness of a topic is assessed based on the clusters that contain topic documents. Rather than selecting all such clusters, we select only enough clusters to jointly cover a given fraction of the topic documents. We refer to this fraction as *Coverage*. All evaluations are replicated at Coverage values of 0.25, 0.50, and 0.75.

Clustering effectiveness at the topic level is characterized using two concepts:

- *Number of selected clusters* (NSC): The number of clusters required to achieve the specified Coverage value.
- *Purity*: The fraction of documents in the selected clusters that belong to the topic.

Each clustering solution, corresponding to a different Resolution value, generates its own Purity and NSC for a given topic. To summarize clustering effectiveness across multiple clustering solutions, we construct a *Purity profile*. A Purity profile represents, for a given topic, the highest Purity value achieved at each NSC value across all clustering solutions, evaluated over a predefined set of NSC values (details in our prior work [5]).

To compare the clustering effectiveness of a topic between the Pure and Baseline networks, we compare the corresponding topic Purity profiles by evaluating their Purity values at the same NSC values. We say that one Purity profile is higher than another if its Purity values are higher for the majority of NSC values.

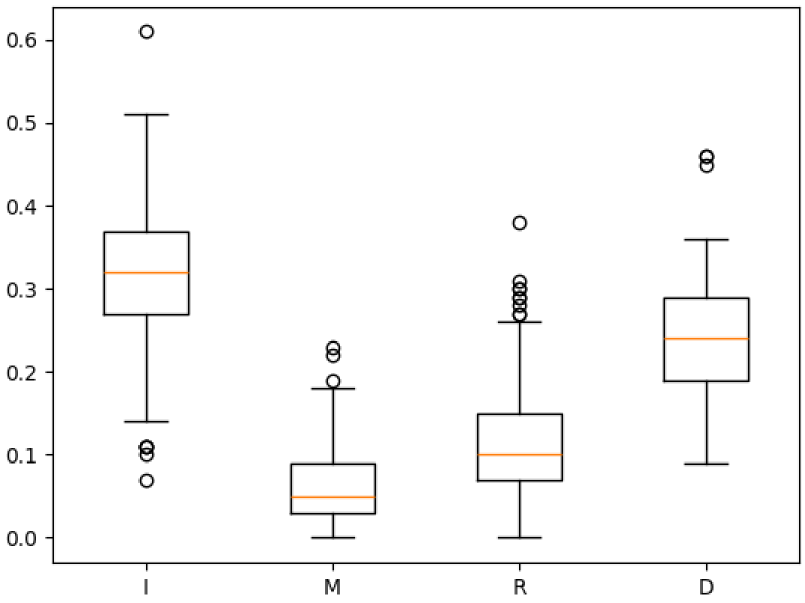

Figure 1: Distribution of the topic categories Purity Difference values among sections

To identify which topic categories perform better in one network than the other, we count how many of their topics had higher clustering effectiveness in one network than the other. We call this value *Purity Difference*, we measure it as the fraction of total topics (i.e. takes value from 0 to 1), and it is one of our key results to answer the research question. A Purity Difference greater than 0.5 indicates that the majority of topics in the category have higher clustering effectiveness in the Pure network than in the Baseline network, and is therefore interpreted as the Pure network outperforming the Baseline network for that topic category.

4. Results

The average Purity Differences per topic category is reported in Table 3, and summarized in Figure 1. For each category, this value is the average of the Purity differences for the three Coverage values.

Across categories, Introduction consistently shows the highest values, followed by Discussion, Results, and Methods. With the exception of two categories in Introduction, no category exceeds a value of 0.5, which indicates that section-specific networks rarely outperform baseline clustering across most topics in any category.

Introduction exhibits elevated values across most categories, with values above 0.30 in the majority of cases. Lower values are mostly observed in categories in the branches Anatomy, Analytical (...), and Phenomena and Processes. Values above 0.35 are mostly observed for categories in the branches Organisms, Psychiatry and Psychology, Disciplines and Occupations, Anthropology (...), Technology (...), Information Science, Named Groups, Health Care, and Geographicals. It achieves values higher than 0.50 on categories [G17] Mathematical Concepts and [H01] Natural Science Discipline.

In Discussion, values are generally above 0.25 for categories in the branches Anatomy, Diseases, Analytical (...), and Geographicals. In Results, above 0.15 for categories in the branches Organisms and Chemicals and Drugs. In Methods, above 0.10 for categories in the branches Disciplines and Occupations, Information Science, Health Care and Geographicals.

5. Discussion

While it is not surprising that only a small number of topic categories outperform the baseline, given that the baseline networks contain more information, it is unexpected that this number is nearly zero. From the perspective of clustering effectiveness, this suggests that there is no systematic advantage to using section-specific citation networks over section-agnostic ones.

We observe that the ordering of sections by Purity Difference matches the ordering by the ratio of citations between the Pure and Baseline networks (Table 2). This ratio reflects the extent to which the Baseline network contains more information than the corresponding Pure network, and suggests that the lower clustering effectiveness of section-specific networks is primarily due to their smaller number of citations. Under this interpretation, section-specific networks might outperform the baseline in some topic categories if they were enriched with additional information. Evidence supporting this possibility can be found in our prior work [5], where networks that combine multiple data sources outperform those based on a single source for several topic categories.

A straightforward way to enrich section-specific networks would be to include citations from all sections while assigning higher weights to citations originating from a specific section. Under this view, the networks used in our experiments can be interpreted as extreme cases: the Baseline network assigns equal weight to all citations with known section origin, while the Pure section network assigns zero weight to citations from other sections. Related approaches have been explored in prior work; for example, Liang et al. [18] assign different weights to citations depending on their section of origin and set the weight of Methods citations to zero. Identifying an optimal weighting scheme between section-specific and non-section citations, however, is likely to pose a substantial computational challenge.

An alternative explanation for the observed correlation with the citation ratio is that the lower performance of section-specific networks is not solely due to reduced information volume, but rather to their increased dissimilarity from the baseline network. Since all citations in the section-specific networks are also present in the baseline network, the correlation may simply reflect how similar each section-specific network is to the baseline.

We also observe consistent differences in which topic categories are most salient across sections. Several branches of topic categories achieve relatively high values in only a single section, such as Psychiatry and Psychology and Anthropology (...) in the Introduction, Anatomy and Analytical (...) in the Discussion, and Chemicals and Drugs in the Results. This is precisely the type of behavior we aimed to uncover: different sections can make different topics more salient in a science map.

We can offer plausible interpretations for why these branches exhibit elevated values in specific sections. In the Results section, documents frequently compare empirical outcomes with prior studies that employed similar experimental setups, which naturally leads to citations involving overlapping chemical substances. In the Introduction section, the branches showing higher values are related to the social sciences. This aligns with the role of the Introduction in biomedical articles in contextualizing research relevance and societal impact. In the Discussion section, the branches showing higher values are related to anatomy and medical techniques. This may reflect the role of this section in research on medical treatments, where authors contrast their treatment with existing approaches, often requiring more detailed reference to affected body parts and the equipment or techniques involved.

Finally, we observe cases in which a topic category achieves high values while the other categories in the branch perform poorly, such as *Archaea [B02]* in the Results section. This highlights the importance of using a more granular system of topic categories, as such patterns would be obscured if analysis were limited to the branch level.

6. Conclusions

We investigated whether restricting citation links to a single document section (Introduction, Methods, Results, or Discussion) changes which topics emerge more cleanly in citation-based science maps. Using large-scale Web of Science citation-section metadata and evaluating clustering effectiveness by

MeSH-based topic categories, we compared section-specific (*Pure*) networks against section-agnostic (*Baseline*) networks that include citations from all sections.

For nearly all topic categories, the majority of topics exhibit lower clustering effectiveness in section-specific citation networks than in section-agnostic ones. Introduction-based maps exhibit the highest performance across all categories, followed by Discussion, Results, and Methods. This ordering of sections matches the ordering by the ratio of citations between the Pure and Baseline networks, suggesting that the cause of the lower performance could be lack of information rather than the sections.

At the same time, the results show that different sections emphasize different kinds of topics. In general terms, the kind of topics that were emphasized only in a single section were: Social sciences in Introduction, parts of the body and medical techniques and equipment in Discussion, and chemical substances in Results. These results indicate that there is a topical difference between science maps derived from sections, although the low performance means that maps made exclusively of sections are inferior in practical terms.

Further research should focus on finding how to take advantage of the topical differences between sections such that high clustering effectiveness can be achieved for these topics. A simple research line is to retain all citations while weighting them by section of origin. Such weighting can be interpreted as interpolating between the two extremes we tested, with uniform weighting in the baseline versus zeroing all but one section in the pure networks. Another research line is to complement section information with the Citation Function Classes of Web of Science [6], which captures the intent of the citation. A final research line is the use of language models optimized to be salient in certain topics. These language models could be trained using only the citations from specific sections, similarly to the work of Liang et al. [18].

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT, Claude in order to: Grammar and spelling check, Paraphrase and reword, Improve writing style. After using this tool, the authors reviewed and edited the content as needed and takes full responsibility for the publication's content.

Table 3: Average Purity Difference per topic category and section. The empty values are topic categories that were removed because they had fewer than 10 topics in that section.

Topic Category	I	M	R	D
[A] Anatomy	0.27	0.05	0.14	0.31
[A01] Body Regions	0.11	0.01	0.06	0.31
[A02] Musculoskeletal System	0.27	0.02	0.07	0.33
[A03] Digestive System	0.23	0.00	0.11	0.36
[A04] Respiratory System	0.11	0.02	0.03	0.24
[A05] Urogenital System	0.11	0.05	0.06	0.26
[A06] Endocrine System	0.07	0.03	0.10	0.26
[A07] Cardiovascular System	0.19	0.04	0.12	0.46
[A08] Nervous System	0.24	0.03	0.07	0.28
[A09] Sense Organs	0.18	0.02	0.11	0.29
[A10] Tissues	0.32	0.03	0.13	0.34
[A11] Cells	0.33	0.08	0.25	0.24
[A12] Fluids and Secretions	0.16	0.05	0.06	0.34
[A13] Animal Structures	0.19	0.01	0.10	0.29
[A14] Stomatognathic System	0.14	0.01	0.02	0.27
[A15] Hemic and Immune Systems	0.28	0.09	0.13	0.24
[A16] Embryonic Structures	0.22	0.08	0.18	0.22
[A17] Integumentary System	0.20	0.00	0.13	0.18
[A18] Plant Structures	0.30	0.00	0.01	0.11
[A19] Fungal Structures	-	-	-	-

Continued on next page

Topic Category	I	M	R	D
[A20] Bacterial Structures	-	-	0.10	-
[A21] Viral Structures	-	-	-	-
[B] Organisms	0.45	0.07	0.18	0.19
[B01] Eukaryota	0.48	0.09	0.17	0.18
[B02] Archaea	0.37	0.09	0.38	0.21
[B03] Bacteria	0.40	0.02	0.19	0.21
[B04] Viruses	0.37	0.03	0.13	0.22
[B05] Organism Forms	0.20	0.09	0.12	0.17
[C] Diseases	0.31	0.04	0.09	0.29
[C01] Infections	0.31	0.02	0.07	0.29
[C04] Neoplasms	0.28	0.07	0.11	0.34
[C05] Musculoskeletal Diseases	0.29	0.05	0.14	0.26
[C06] Digestive System Diseases	0.30	0.07	0.16	0.29
[C07] Stomatognathic Diseases	0.30	0.04	0.07	0.31
[C08] Respiratory Tract Diseases	0.29	0.05	0.06	0.27
[C09] Otorhinolaryngologic Diseases	0.36	0.04	0.09	0.28
[C10] Nervous System Diseases	0.36	0.04	0.09	0.26
[C11] Eye Diseases	0.30	0.04	0.11	0.29
[C12] Urogenital Diseases	0.26	0.03	0.07	0.31
[C14] Cardiovascular Diseases	0.33	0.05	0.11	0.34
[C15] Hemic and Lymphatic Diseases	0.33	0.05	0.08	0.30
[C16] Congenital, Hereditary, and Neonatal Diseases and Abnormalities	0.36	0.04	0.09	0.24
[C17] Skin and Connective Tissue Diseases	0.30	0.04	0.08	0.29
[C18] Nutritional and Metabolic Diseases	0.35	0.03	0.05	0.24
[C19] Endocrine System Diseases	0.25	0.02	0.10	0.31
[C20] Immune System Diseases	0.37	0.06	0.07	0.33
[C21] Disorders of Environmental Origin	-	-	-	-
[C22] Animal Diseases	0.34	0.02	0.09	0.21
[C23] Pathological Conditions, Signs and Symptoms	0.31	0.05	0.09	0.29
[C24] Occupational Diseases	-	-	-	-
[C25] Chemically-Induced Disorders	0.29	0.01	0.07	0.22
[C26] Wounds and Injuries	0.27	0.03	0.09	0.31
[D] Chemicals and Drugs	0.34	0.10	0.27	0.27
[D01] Inorganic Chemicals	0.35	0.08	0.22	0.16
[D02] Organic Chemicals	0.43	0.10	0.28	0.23
[D03] Heterocyclic Compounds	0.40	0.10	0.23	0.26
[D04] Polycyclic Compounds	0.42	0.12	0.20	0.29
[D05] Macromolecular Substances	0.35	0.08	0.29	0.33
[D06] Hormones, Hormone Substitutes, and Hormone Antagonists	0.32	0.04	0.13	0.34
[D08] Enzymes and Coenzymes	0.30	0.10	0.30	0.27
[D09] Carbohydrates	0.31	0.12	0.29	0.24
[D10] Lipids	0.39	0.11	0.25	0.20
[D12] Amino Acids, Peptides, and Proteins	0.30	0.09	0.30	0.32
[D13] Nucleic Acids, Nucleotides, and Nucleosides	0.42	0.18	0.31	0.22
[D20] Complex Mixtures	0.38	0.04	0.15	0.21
[D23] Biological Factors	0.28	0.11	0.24	0.28
[D25] Biomedical and Dental Materials	0.24	0.05	0.07	0.19
[D26] Pharmaceutical Preparations	0.32	0.08	0.06	0.28
[D27] Chemical Actions and Uses	0.34	0.09	0.16	0.23
[E] Analytical, Diagnostic and Therapeutic Techniques, and Equipment	0.30	0.09	0.13	0.32
[E01] Diagnosis	0.30	0.11	0.13	0.32
[E02] Therapeutics	0.26	0.05	0.11	0.32
[E03] Anesthesia and Analgesia	0.10	0.02	0.02	0.19
[E04] Surgical Procedures, Operative	0.24	0.05	0.16	0.46
[E05] Investigative Techniques	0.36	0.16	0.15	0.22

Continued on next page

Topic Category	I	M	R	D
[E06] Dentistry	0.28	0.05	0.09	0.45
[E07] Equipment and Supplies	0.38	0.04	0.08	0.31
[F] Psychiatry and Psychology	0.38	0.02	0.04	0.16
[F01] Behavior and Behavior Mechanisms	0.40	0.02	0.03	0.14
[F02] Psychological Phenomena	0.37	0.01	0.05	0.17
[F03] Mental Disorders	0.42	0.03	0.04	0.20
[F04] Behavioral Disciplines and Activities	0.33	0.04	0.05	0.15
[G] Phenomena and Processes	0.34	0.08	0.17	0.23
[G01] Physical Phenomena	0.40	0.04	0.07	0.10
[G02] Chemical Phenomena	0.45	0.19	0.26	0.10
[G03] Metabolism	0.33	0.15	0.20	0.12
[G04] Cell Physiological Phenomena	0.22	0.09	0.18	0.15
[G05] Genetic Phenomena	0.40	0.12	0.27	0.28
[G06] Microbiological Phenomena	0.26	0.00	0.07	0.34
[G07] Physiological Phenomena	0.32	0.07	0.13	0.22
[G08] Reproductive and Urinary Physiological Phenomena	0.28	0.01	0.03	0.23
[G09] Circulatory and Respiratory Physiological Phenomena	0.20	0.04	0.10	0.26
[G10] Digestive System and Oral Physiological Phenomena	0.19	0.02	0.02	0.19
[G11] Musculoskeletal and Neural Physiological Phenomena	0.22	0.04	0.03	0.24
[G12] Immune System Phenomena	0.29	0.08	0.17	0.27
[G13] Integumentary System Physiological Phenomena	-	-	-	-
[G14] Ocular Physiological Phenomena	0.29	0.00	0.10	0.31
[G15] Plant Physiological Phenomena	0.29	0.00	0.08	0.23
[G16] Biological Phenomena	0.46	0.07	0.05	0.18
[G17] Mathematical Concepts	0.61	0.18	0.00	0.14
[H] Disciplines and Occupations	0.41	0.12	0.11	0.20
[H01] Natural Science Disciplines	0.51	0.16	0.14	0.14
[H02] Health Occupations	0.33	0.10	0.08	0.23
[I] Anthropology, Education, Sociology, and Social Phenomena	0.39	0.06	0.07	0.17
[I01] Social Sciences	0.42	0.07	0.09	0.13
[I02] Education	0.26	0.01	0.11	0.14
[I03] Human Activities	0.30	0.03	0.05	0.25
[J] Technology, Industry, and Agriculture	0.41	0.05	0.13	0.15
[J01] Technology, Industry, and Agriculture	0.41	0.06	0.12	0.16
[J02] Food and Beverages	0.26	0.03	0.12	0.10
[J03] Non-Medical Public and Private Facilities	0.30	0.03	0.04	0.11
[K] Humanities	0.35	0.09	0.06	0.11
[L] Information Science	0.44	0.22	0.15	0.16
[M] Named Groups	0.36	0.06	0.06	0.24
[N] Health Care	0.36	0.10	0.09	0.20
[N01] Population Characteristics	0.35	0.14	0.06	0.09
[N02] Health Care Facilities Workforce and Services	0.30	0.05	0.09	0.26
[N03] Health Care Economics and Organizations	0.39	0.13	0.05	0.15
[N04] Health Services Administration	0.36	0.12	0.12	0.19
[N05] Health Care Quality, Access, and Evaluation	0.32	0.23	0.13	0.23
[N06] Environment and Public Health	0.41	0.12	0.07	0.17
[Z] Geographicals	0.42	0.18	0.11	0.32

References

[1] M. Zitt, Meso-level retrieval: IR-bibliometrics interplay and hybrid citation-words methods in scientific fields delineation, Scientometrics 102 (2015) 2223–2245. doi:10.1007/s11192-014-1482-5.
[2] J. P. Bascur, S. Verberne, N. J. van Eck, L. Waltman, Academic information retrieval using citation

clusters: In-depth evaluation based on systematic reviews, Scientometrics 128 (2023) 2895–2921. doi:10.1007/S11192-023-04681-X.

[3] M. Yuan, J. Zobel, P. Lin, Measurement of clustering effectiveness for document collections, Information Retrieval Journal 25 (2022) 239–268. URL: https://doi.org/10.1007/s10791-021-09401-8. doi:10.1007/s10791-021-09401-8.

[4] J. P. Bascur, S. Verberne, N. J. van Eck, L. Waltman, Which topics are best represented by science maps? an analysis of clustering effectiveness for citation and text similarity networks, Scientometrics 130 (2025) 1181–1199. URL: https://doi.org/10.1007/s11192-024-05218-6. doi:10.1007/s11192-024-05218-6.

[5] J. P. Bascur, R. Costas, S. Verberne, Use of diverse data sources to control which topics emerge in a science map, 2025. URL: https://arxiv.org/abs/2412.07550. arXiv:2412.07550.

[6] Enriched cited references - citation function class, https://webofscience.zendesk.com/hc/en-us/articles/27713957430033-Enriched-Cited-References-Citation-Function-Class, 2025. Web of Science Help Center, Clarivate. Accessed: 2026-01-16.

[7] G. Lin, N. J. van Eck, H. Hou, Z. Hu, The changing role of cited papers over time: An analysis of highly cited papers based on a large full-text dataset, 2025. URL: https://arxiv.org/abs/2509.04190. arXiv:2509.04190.

[8] K. W. Boyack, N. J. van Eck, G. Colavizza, L. Waltman, Characterizing in-text citations in scientific articles: A large-scale analysis, Journal of Informetrics 12 (2018) 59–73. URL: https://www.sciencedirect.com/science/article/pii/S1751157717303516. doi:https://doi.org/10.1016/j.joi.2017.11.005.

[9] M. Thelwall, Should citations be counted separately from each originating section?, Journal of Informetrics 13 (2019) 658–678. URL: https://www.sciencedirect.com/science/article/pii/S1751157718305121. doi:https://doi.org/10.1016/j.joi.2019.03.009.

[10] Z. Hu, C. Chen, Z. Liu, Where are citations located in the body of scientific articles? a study of the distributions of citation locations, Journal of Informetrics 7 (2013) 887–896. URL: https://www.sciencedirect.com/science/article/pii/S1751157713000692. doi:https://doi.org/10.1016/j.joi.2013.08.005.

[11] M. Thelwall, Are classic references cited first? an analysis of citation order within article sections, Scientometrics 120 (2019) 723–731. doi:10.1007/s11192-019-03142-8.

[12] C. Lu, Y. Ding, C. Zhang, Understanding the impact change of a highly cited article: a content-based citation analysis, Scientometrics 112 (2017) 927–945. URL: https://doi.org/10.1007/s11192-017-2398-7. doi:10.1007/s11192-017-2398-7.

[13] W. Otto, B. Ghavimi, P. Mayr, R. Piryani, V. K. Singh, Highly cited references in plos one and their in-text usage over time, 2019. URL: https://arxiv.org/abs/1903.11693. arXiv:1903.11693.

[14] J. H. Jeon, J. J. Jung, Article ranking with location-based weight in contextual citation network, Journal of Informetrics 18 (2024) 101591. URL: https://www.sciencedirect.com/science/article/pii/S1751157724001032. doi:https://doi.org/10.1016/j.joi.2024.101591.

[15] P. Donner, S. Stahlschmidt, R. Haunschild, L. Bornmann, Are citation context information stronger related to peer ratings than citation counts? a descriptive analysis, in: Proceedings of the 20th International Conference on Scientometrics & Informetrics, volume 1, 2025, pp. 294–301. doi:10.51408/issi2025_049.

[16] M. A. Abbas, S. Ajayi, M. Bilal, A. Oyegoke, M. Pasha, H. T. Ali, A deep learning approach for context-aware citation recommendation using rhetorical zone classification and similarity to overcome cold-start problem, Journal of Ambient Intelligence and Humanized Computing 15 (2024) 419–433. doi:10.1007/s12652-022-03899-6.

[17] Y. Zhang, Q. Ma, Dual attention model for citation recommendation, in: D. Scott, N. Bel, C. Zong (Eds.), Proceedings of the 28th International Conference on Computational Linguistics, International Committee on Computational Linguistics, Barcelona, Spain (Online), 2020, pp. 3179–3189. URL: https://aclanthology.org/2020.coling-main.283/. doi:10.18653/v1/2020.coling-main.283.

[18] Z. Liang, N. J. van Eck, X. Wu, J. Mao, G. Li, Citation importance-aware document representation

learning for large-scale science mapping, Information Processing & Management 63 (2026) 104557. URL: https://www.sciencedirect.com/science/article/pii/S0306457325004984. doi:https://doi.org/10.1016/j.ipm.2025.104557.
[19] L. Waltman, K. W. Boyack, G. Colavizza, N. J. van Eck, A principled methodology for comparing relatedness measures for clustering publications, Quantitative Science Studies 1 (2020) 691–713. doi:10.1162/qss_a_00035.
[20] T.-K. Hsiao, V. I. Torvik, Opcitance: Citation contexts identified from the pubmed central open access articles, Scientific Data 10 (2023) 243. doi:10.1038/s41597-023-02134-x.
[21] V. A. Traag, L. Waltman, N. J. Van Eck, From louvain to leiden: guaranteeing well-connected communities, Scientific reports 9 (2019) 1–12. doi:10.1038/s41598-019-41695-z.
[22] G. Csardi, T. Nepusz, The igraph software package for complex network research, InterJournal, Complex Systems 1695 (2006).