Algorithms for Scientific Documents: Past and Present

Jesús M. Álvarez-Llorente; Vicente P. Guerrero-Bote; Félix De-Moya-Anegón

Cómo citar este artículo:

Álvarez-Llorente, Jesús M.; **Guerrero-Bote, Vicente P.**; **De-Moya-Anegón, Félix** (2025). "Algorithms for Scientific Documents: Past and Present [Algoritmos para documentos científicos: pasado y presente]". *Infonomy*, 3(4) e25026.
https://doi.org/10.3145/infonomy.25.026

Received: 01-08-2025
Approved: 10-09-2025

Jesús M. Álvarez-Llorente
https://orcid.org/0000-0002-4901-3457
https://directorioexit.info/ficha6849
Universidad de Extremadura
Departamento de Ingeniería de Sistemas Informáticos y Telemáticos
Plazuela Ibn Marwam
06071 Badajoz, España
llorente@unex.es

Vicente P. Guerrero-Bote
https://orcid.org/0000-0003-4821-9768
https://directorioexit.info/ficha596
Universidad de Extremadura
Departamento de Ingeniería de Sistemas Informáticos y Telemáticos
Plazuela Ibn Marwam
06071 Badajoz, España
guerrero@unex.es

Félix De-Moya-Anegón
https://orcid.org/0000-0002-0255-8628
https://directorioexit.info/ficha92
SCImago Research Group
18220 Granada, España
felix.moya@scimago.es

Abstract
This study offers a comprehensive overview of document-level classification algorithms in scientific research, proposed as an alternative to the journal-based categorizations employed by major bibliographic databases such as *Web of Science* and *Scopus*. These journal-driven schemes often introduce significant inaccuracies in both information retrieval and research evaluation, as they fail to categorize articles in accordance with their actual content. First, we provide a historical review of the main approaches developed since the emergence of scientific databases, highlighting their contributions as well as their limitations. Automatic clustering techniques and community detection algorithms have represented important advances in the organization of scientific knowledge, yet they cannot serve as a practical substitute for journal-based classifications. Other approaches, such as those relying on neural networks or text mining, face scalability issues that prevent their application at the global level of science. The most recent and promising strategies are built upon simple algorithms that, starting from existing journal categorizations, reclassify articles into the same thematic hierarchies used by bibliographic databases, relying primarily on the analysis of straightforward citation and reference patterns.

Keywords
Classification algorithms; Document-level classifications; Classifications; Science classification; Scientific databases; Scientometrics; Citation; Classification schemes; *ASJC*; *Scopus*; *Web of Science*.

1. Introduction
"Science is not science if it is not shared" is a phrase frequently cited today, attributed to Nathan Robinson, a researcher at the *Institut de Ciències del Mar* (*CSIC*). Although Robinson originally used it to highlight the importance of science communication through modern channels —particularly social media— it is clear that the progress of science would be impossible without the traditional exchange of knowledge among researchers through established publication mechanisms. These include books and conference proceedings, but especially scientific journals, which in turn serve as the main vehicle for the dissemination of articles. Yet, dissemination does not end with the act of publishing: other researchers must be able to locate and access this published information. In a globalized world with a colossal annual volume of scientific output, large bibliographic databases such as *Scopus* and *Web of Science* (*WoS*) —which index this output, primarily journal articles, and make it searchable for researchers worldwide— have become indispensable tools for the advancement of science.

Within these databases, it is necessary to organize publications into scientific categories. The most traditional method is to classify journals according to their thematic scope. In both information retrieval and scientometric studies, it is common practice to extend these journal-level categories to all articles published within them, regardless of whether the actual content of a given article fully aligns with the journal's thematic focus. This approach introduces significant problems for information retrieval, due to the high level of inaccuracy it generates, and poses even greater challenges for scientometrics. Quantifying research output by discipline, as well as measuring scientific impact —an essential factor in research funding— becomes problematic, since citation practices vary widely across disciplines.

This limitation has been recognized since the very inception of bibliographic databases, leading to numerous attempts to classify articles individually according to their own characteristics, rather than by the journals in which they appear. These approaches are often referred to as *document-level classifications* or *item-by-item classifications*. Depending on their design and intended purpose, two broad strategies can be identified.

The first involves the use of automatic clustering techniques or community detection algorithms, which group publications with similar characteristics into clusters or communities —categories not predefined but emerging from the data itself—. The second strategy assigns each publication to one of the thematic categories of a pre-established classification scheme —often the same used by the database for journal-level classification— through a process of reclustering or, more accurately, (re)classification. The assignment is typically based on the analysis of citation networks (citation, co-citation, or bibliographic coupling), text mining techniques (mainly term frequency), or hybrid methods that combine both to improve accuracy.

Traditional implementations of these methods have faced significant limitations. Clustering techniques, for example, suffer from stability issues, often producing different results with each run. They also tend to yield classification schemes that diverge considerably from those currently employed by bibliographic databases, introducing inconsistency. Moreover, they are computationally demanding, which restricts their application to small subsets of scientific output, and many do not allow publications to belong to multiple categories simultaneously —an unrealistic constraint—. These same limitations are also encountered in (re)classification methods, particularly when they incorporate text mining or the more complex forms of citation network analysis such as co-citation and bibliographic coupling.

In short, there is a clear need for a system of individual article classification that improves upon journal-based categorizations while overcoming known limitations: one that is stable, consistent with the classification schemes already accepted by the scientific community, scalable to the entirety of science, flexible enough to assign articles to multiple categories, and sensitive to disciplinary differences —an essential condition for any classification system to be genuinely useful in today's scientific landscape—.

When referring to these methods, given the scale of information they aim to process, it is appropriate to describe them as *algorithms*. In the current context of artificial intelligence, however, the term “algorithm” has acquired an undeservedly negative connotation. It is often portrayed as an uncontrollable and opaque force that monitors and dictates our lives, even threatening their very existence. Yet, in reality, an algorithm is nothing more than a sequence of instructions designed to accomplish a task —something as simple and harmless as a recipe—. Computers, which are responsible for automatically processing the vast amounts of information we handle, operate precisely through algorithms. An algorithm is simply a formal description of how to achieve a desired outcome.

This article provides a retrospective analysis of the main document-level classification algorithms proposed in recent decades and examines how they have paved the way for the most recent and promising approaches, which are discussed in greater detail in the following sections.

2. Background

Web of Science (*WoS*) and *Scopus* are unquestionably the two most widely accepted scientific databases within the research community. Both classify scientific journals into categories. In practice, for purposes of information retrieval as well as in many bibliometric studies, articles inherit the categories assigned to the journal in which they were published. There are, however, important differences between the classification schemes of the two databases.

Scopus employs the *All Science Journal Classification* (*ASJC*) scheme (**Gómez-Crisóstomo**, 2011; **Wang; Waltman**, 2016). It consists of 27 broad subject areas, one of which is explicitly multidisciplinary (“1000 Multidisciplinary”), where journals such as *Nature* or *Science* are placed. The remaining 26 subject areas are subdivided into 311 more specific subject categories. Notably, each of these 26 areas includes a “miscellaneous” category, designed to accommodate journals covering diverse topics within the area or those that cannot be clearly assigned to a specific category.

WoS, by contrast, establishes five broad research areas at its highest level (Arts and Humanities, Life Sciences and Biomedicine, Physical Sciences, Social Sciences, and Technology). Within these five areas, a second level distinguishes —currently— a total of 254 subject categories or disciplines. Unlike *Scopus*, *WoS* does not define a dedicated multidisciplinary area, but it does include a pure multidisciplinary category, as well as several others explicitly designated as multidisciplinary or interdisciplinary (e.g., *Computer Science, Interdisciplinary Applications* or *Humanities, Multidisciplinary*). Both schemes are widely recognized and extensively used in studies on the structure of science (**Leydesdorff** *et al.*, 2010; 2015; **Hassan-Montero** *et al.*, 2014). However, they are not the only classification systems available. **Gläser** *et al.* (2017) provide a comprehensive overview of various systems proposed throughout history. Among their conclusions, they stress that one of the most persistent challenges faced by all classification systems lies in defining research areas, given that research fronts evolve continuously and boundaries between categories are often blurred.

In both *WoS* and *Scopus* (as well as in many other bibliographic databases), journals can be assigned to multiple categories. By contrast, it is relatively rare to find articles that truly span multiple subjects. This situation underscores the necessity of multidisciplinary categories, which serve to accommodate numerous journals with very broad thematic scopes. In fact, **Zhang** & **Shen** (2024) argue that many journals not officially designated as multidisciplinary should arguably be classified as such, given the diversity of topics they actually cover.

According to **Wang** & **Waltman** (2016), the average number of categories per journal is 1.6 in *WoS* and 2.1 in *Scopus*. When the calculation is made at the article level (i.e., by extending journal categories to all the articles they publish), the average for *Scopus* rises to 2.5, with a clear upward trend over time. Importantly, not all articles in a given journal necessarily fall into each of the categories assigned to that journal. Numerous studies point out the mismatch between the categories attributed to journals by databases and the actual topics of their articles. **Thelwall** & **Pinfield** (2024), for example, emphasize that such imprecision occurs especially in journals lacking a clearly specialized profile, particularly those assigned to multidisciplinary or miscellaneous categories.

This lack of precision undermines the reliability of journal-based classifications, particularly in the normalization of citation indicators. Differences in publication and citation habits across disciplines make normalization essential, yet imprecise classifications hinder its effectiveness, thereby compromising crucial aspects of research evaluation and funding. Several studies have highlighted this issue:

- **Althouse** *et al.* (2009) analyzed disciplinary differences, their evolution over time, and their effect on the impact factor.
- **Opthof** & **Leydesdorff** (2010) warned of the consequences for the scientific evaluation of researchers.
- **Lancho-Barrantes** *et al.* (2010b) demonstrated that impact factors in major databases—one of the main indicators used to evaluate researchers—were highly correlated with the average number of active references (also indexed in the database) per article during the corresponding period, underscoring the need for normalization measures.
- **Guerrero-Bote** & **De-Moya-Anegón** (2012) proposed improvements to the SJR indicator by adding further normalization adjustments.
- **Bornmann** & **Leydesdorff** (2017) reported significant asymmetries in citation impact across six major disciplines.
- **Bornmann** *et al.* (2019) compared several bibliometric indicators with different normalization techniques.
- **Andersen** (2023) examined statistical risks associated with certain normalization approaches applied across highly heterogeneous disciplines.
- **Thelwall** & **Pinfield** (2024) highlighted search and indicator calculation problems arising in journals assigned to multidisciplinary or miscellaneous categories.

The problem of normalization becomes particularly complex in the case of journals classified as multidisciplinary or miscellaneous. Articles published in such journals, when assigned journal-level categories, risk being left without a specific thematic profile —or conversely, appearing to cover an excessive range of subjects—. For this reason, much research has focused on developing methods to classify articles published in multidisciplinary journals (**Glänzel** *et al.*, 1999a, 1999b, 2021; **Fang**, 2015; **Zhang** *et al.*, 2022; **Zhang**; **Shen**, 2024).

At the opposite extreme lies the approach of assigning each article to a single category, a principle that holds particular scientometric interest in certain cases. Proposals such as those by **Milojević** (2020) or **Waltman** & **Van Eck** (2012) focus specifically on this strategy. Nevertheless, most recent approaches allow multiple assignments (e.g., **Fang**, 2015; **Glänzel** *et al.*, 2021; **Zhang** *et al.*, 2022). Zhang *et al.* (2022) explicitly argue that multiple assignments are essential, since authors themselves often conceptualize their work as spanning more than one field. Similarly, **Zhang** *et al.* (2016), **Huang** *et al.* (2021), and **Thijs** *et al.* (2021) present evidence of the inherently multidisciplinary nature of science.

That said, assigning articles to an excessive number of categories —particularly when the connection to some categories is weak— merely reproduces the imprecision of journal-based classification, precisely the problem these efforts seek to overcome. The literature offers several approaches to limiting the number of categories:

- **Fang** (2015) proposed applying a threshold to discard weakly related categories.
- **Waltman** *et al.* (2020) restricted assignments to the top *N* categories with the strongest relationships.
- **Glänzel** *et al.* (2021) developed a more nuanced criterion, allowing only those assignments whose strength is not significantly lower than that of the strongest category. For instance, if a paper has a 90% association with category A and 10% with B, only A is retained; but if the association is 55% with A and 45% with B, both are accepted.

In addition to limiting the number of categories, **Glänzel** *et al.* (2021) further stipulated that no assignments should be made to general multidisciplinary categories, an idea also supported by **Milojević** (2020). At the article level, this restriction is logical: while a paper may cover several subjects, even many, it cannot possibly encompass all scientific fields simultaneously.

3. Document-Level Classification Algorithms in Historical Perspective

Since the late twentieth century, the classification of scientific documents has been approached through a wide range of methodological strategies. One of the earliest antecedents can be traced to thesaurus-generation algorithms for knowledge organization (**Rees-Potter**, 1989). Subsequent developments explored algorithms based on citation links between scientific journals (**Marshakova-Shaikevich**, 2005; **De-Moya-Anegón** *et al.*, 2006; **Schildt** *et al.*, 2006), as well as classification methods applied to specific types of documents such as patents (**Lai**; **Wu**, 2005) or even entire journals (**Zhang** *et al.*, 2010).

In parallel, automatic clustering and community detection algorithms were developed, notably those proposed by **Clauset** *et al.* (2004) and **Blondel** *et al.* (2008), which enabled the identification of emerging thematic structures in large volumes of scientific literature. However, these approaches also present important limitations: the incorporation of new documents can significantly alter results, and the inherent randomness of some algorithms may lead to inconsistent classifications even with identical input data (**Klavans**; **Boyack**, 2005, 2006; **Waltman**; **Van Eck**, 2012; **Janssens** *et al.*, 2008; 2009).

Within this context, neural network–based algorithms began to show considerable potential for document organization. Early studies (**Guerrero-Bote** *et al.*, 2002) demonstrated their capacity to detect meaningful patterns in titles, abstracts, and full texts. These techniques have been applied in both supervised learning contexts (**Eykens** *et al.*, 2019) and unsupervised settings (**Kandimalla** *et al.*, 2021), the latter being particularly dependent on citation networks as the basis for self-learning.

As these approaches matured, hybrid methods emerged, combining citation-based relationships with textual analysis in order to improve classification accuracy. These integrated methodologies have been widely studied (**Glenisson** *et al.*, 2005; Janssens *et al.*, 2006; 2008; 2009; **Boyack** *et al.*, 2013; **Boyack**; **Klavans**, 2020), benefiting from advances in content analysis techniques (**Boyack** *et al.*, 2011), which extend from titles and keywords to full-text semantic analysis.

With respect to citation relationships, three fundamental mechanisms have been employed: simple citation, which links a publication to another when it is cited in its references; co-citation, which connects two documents when both are cited by a third; and bibliographic coupling, which links two works that share one or more common references. A detailed review of these techniques is provided in **Šubelj** *et al.* (2016).

The conceptual simplicity of direct citation makes it especially useful for handling large volumes of publications with relatively low computational cost. This advantage has favored its use in large-scale studies such as those by **Boyack** & **Klavans** (2010), **Waltman** & **Van Eck** (2012), and **Klavans** & **Boyack** (2006; 2016). However, one must also acknowledge that some publications contain no references, while many others include very few. For example, up to 6.5% of *Scopus*-indexed publications in 2020 contained no active references (i.e., references indexed in the database), and 10.4% contained fewer than three (**Álvarez-Llorente** *et al.*, 2024). Such cases either prevent or significantly complicate classification based solely on citation data without additional analytical effort.

Even so, hybrid algorithms have generally demonstrated superior performance in terms of both accuracy and coherence. **Boyack** & **Klavans** (2020) concluded that combining citation relationships with textual analysis outperforms not only the isolated application of each technique but also other approaches based solely on citation relationships.

Not all studies, however, support this view. **Chumachenko** *et al.* (2022) warned that textual analysis may introduce noise due to the recurrence of common phrases across multiple disciplines. Their proposal focused on semantic full-text analysis, extracting relevant concepts and using entropy measures to distinguish between core concepts and generic terms. Similarly, **Sachini** *et al.* (2022) applied neural network algorithms trained on text to (re)classify publications in the field of Artificial Intelligence. While their results were promising, they also emphasized the high computational costs of such methods, which limit scalability to very large datasets.

Clustering and other algorithms that generate classification schemes automatically also tend to produce categories that diverge considerably from those adopted by bibliographic databases, often necessitating manual intervention by experts. To support this process, visualization tools such as *VOSviewer* (**Van Eck**; **Waltman**, 2010) and *SCImago Graphica* (**Hassan-Montero** *et al.*, 2022) have been developed, allowing interactive exploration of thematic communities and information flows.

To mitigate the variability introduced by clustering and neural network algorithms — which typically incorporate random factors into their logic— and to improve classification coherence, some recent studies have chosen to rely on the thematic schemes already used by major bibliographic databases. A prominent example, to be discussed in the next section, is the work of **Milojević** (2020), who assigns *WoS* categories to individual articles based on their cited references. The rationale is straightforward: if most of a document's references belong to journals classified in a given category, the article itself is assigned to that same category.

Along similar lines, **Glänzel** *et al.* (2021) proposed a document-level classification algorithm known as the "Multi-Generation Parametric Model," which integrates many of the approaches outlined above, along with additional innovative concepts. More recently, two algorithms inspired by **Glänzel** *et al.* (2021) but incorporating new advances —*M3-AWC-0.8* (**Álvarez-Llorente** *et al.*, 2024) and *U1-F-0.8* (**Álvarez-Llorente** *et al.*, 2025)— have been introduced. These will be discussed in greater detail in the following section.

As a synthesis, Figure 1 presents a timeline of the algorithms discussed here, with data partly drawn from **Álvarez-Llorente** (2025). Algorithms that will be analyzed in greater depth in the next section are highlighted in yellow.

The proliferation of algorithms based on diverse methodological foundations and yielding often divergent results highlights the need for robust comparative criteria to assess their relative quality. **Waltman** *et al.* (2020) argue that any classification system may be considered valid provided it aligns with its intended purpose, while also stressing the importance of objective accuracy metrics and mechanisms for comparing classification schemes. To enable meaningful comparisons, they propose employing a third classification as a benchmark, with a methodology substantially different from the two under analysis. For example, when comparing clusterings based on citation relationships, they suggest using a classification derived from textual analysis as the evaluation criterion, and vice versa.

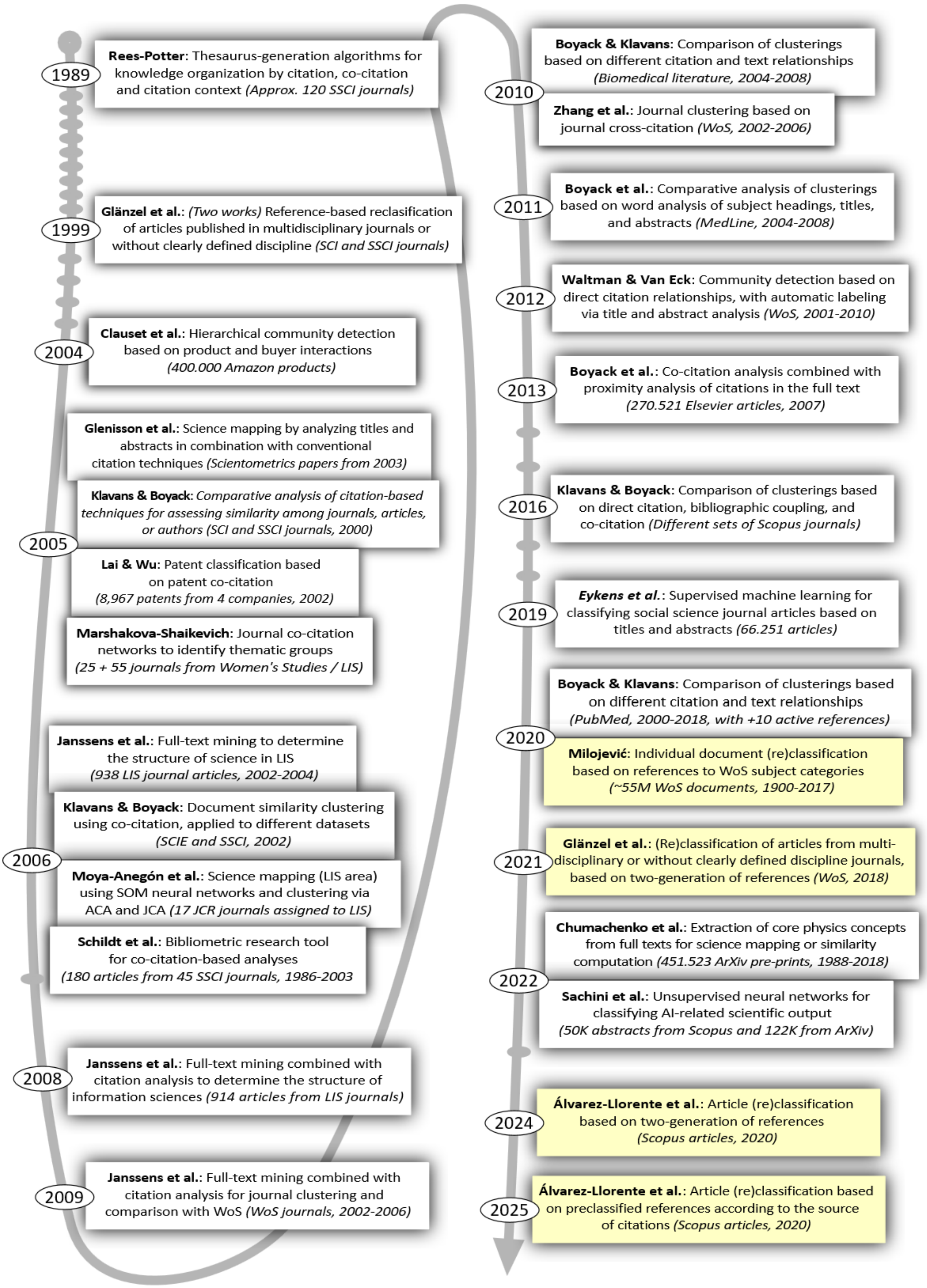

Figure 1. Timeline of document-level classification algorithms addressed in this study.

In line with this comparative approach, **Boyack** & **Klavans** (2020) developed an evaluation model incorporating three distinct types of accuracy metrics: one based on the references of highly cited documents (those with more than 100 citations), another grounded in textual analysis, and a third considering pairs of documents linked by funding acknowledgments. Their results confirmed that combining citation data with textual content analysis produces superior outcomes —not only compared to the isolated use of each technique, but also relative to other citation-only approaches— despite the substantial computational costs associated with such hybrid methods.

When evaluating classification accuracy, however, it seems intuitive to consider manual classifications by experts —or better still, by the authors themselves— as the most reliable benchmark. The underlying assumption is that manual classification, by its radically different nature, offers a valuable perspective against which to assess automated systems.

One of the first attempts in this vein was that of **Šubelj** *et al.* (2016), who complemented quantitative indicators with qualitative assessments by scientometric experts to evaluate the quality of automatic clusterings. Another example is the study by **Milojević** (2020), in which authors were asked to choose, for their own publications, whether the traditional journal-based classification under *WoS* or the newly proposed document-level scheme was more appropriate. The sample included 142 articles, all assigned to a single category without a multidisciplinary character. A second phase involved manual classification of an additional 100 publications, in which authors were asked to assign a category based solely on the article's title and abstract.

Similarly, **Zhang** *et al.* (2022) conducted a study focused on publications in *Nature*, comparing three classification systems: thematic labels assigned by authors upon submission (based on the journal's thematic hierarchy, **McGillivray** & **Astell**, 2019), the *WoS* classification via *InCites* (based on cited references), and the Fields of Research (FoR) scheme from the *Dimensions* database, which employs machine learning techniques. Although the study offered valuable comparisons, its main limitation lay in its restricted applicability to a single editorial environment.

These studies underscore both the value and the limitations of manual classifications. On the one hand, they provide a human perspective, arguably closer to authors' original intentions, against which to benchmark automated systems. On the other hand, they face clear scalability constraints, as their manual nature precludes large-scale application. Moreover, author-provided labels inevitably introduce a degree of subjectivity, leading to inconsistencies and discrepancies with classifications based on references or textual content (**Shu** *et al.*, 2019; **Zhang** *et al.*, 2022).

Despite these limitations, author input remains a valuable resource for assessing the validity of thematic classifications. To become a truly useful tool, however, it requires a broad and representative corpus of publications labeled by their own authors, reflecting adequate thematic, geographic, and disciplinary diversity.

With this goal, **Álvarez-Llorente** *et al.* (2023) introduced the *Author's Assignation Collection* (*AAC*), the largest dataset to date of publications categorized by their corresponding authors. This collection comprises over 14,000 articles across diverse fields of knowledge and regions, classified according to a "fractional *ASJC*" scheme derived from *Scopus*' *ASJC* system. In line with the methodological principles outlined in the Background section, the scheme excludes the Multidisciplinary area and miscellaneous categories, and allows authors to assign their work —on a weighted basis— to up to five categories. Results confirmed significant discrepancies between manual classifications and those generated automatically through references or textual content. Nevertheless, manual classification is not intended as an operational alternative for organizing science, but rather as a validation and benchmarking tool, offering a qualitatively distinct reference point especially valuable for evaluating the coherence and precision of automated systems.

Author's Assignation Collection (*AAC*) is the largest dataset to date of publications categorized by their corresponding authors

4. Document-Level Classification Algorithms: The Present

The historical analysis of document-level classification algorithms has led to the thesis that a strategy based primarily on citation relationships, complemented by pre-established classification schemes, offers a balanced and viable solution for large-scale thematic classification of scientific output. In this section, we present four recent proposals that share these characteristics, accompanied by a brief analysis of their algorithms. Their main features are summarized in Table 1.

5. Practical Method to Reclassify *WoS* Articles

This proposal, introduced in the article *"Practical method to reclassify WoS articles into unique subject categories and broad disciplines"* (**Milojević**, 2020), hereafter referred to as the Milojević proposal, provides a method to (re)classify *WoS* publications at the individual level. It relies on the *WoS* classification scheme (excluding interdisciplinary categories) as well as the 14 broad areas defined in the *NSF WebCASPAR Broad Field* (**Javitz** *et al.*, 2010), forcing assignments to a single discipline in both cases.

The method is based on "classifying references," i.e., references to journals assigned to a single category (excluding multidisciplinary ones). Each publication is assigned to the category that accounts for the majority of its classifying references. It applies to all *WoS* publications up to 2017 that contain at least one reference to another *WoS* item, although many documents remain unclassified due to a lack of such references. Iterative applications of the method increase the number of classified articles.

Identification	Practical method to reclassify *WoS* articles	Multi-Generation Parametric Model	*M3-AWC-0.8*	*U1-F-0.8*
Reference (Year)	**Milojević** (2020)	**Glänzel** *et al.* (2021)	**Álvarez-Llorente** *et al.* (2024)	**Álvarez-Llorente** *et al.* (2025)
Scheme	*NSF WebCASPAR Broad Field // WoS*	Modified Leuven-Budapest	Fractional *ASJC*	Fractional *ASJC*
Areas / Catego-ries	14 // 252[1]	15 / 73	26 / 285	26 / 285
Máx. categories per article	1	3	5	5
Relationship type	References	Two-generation references	Two-generation references	References pre-classified by their citers
Dataset	*WoS Core Collection* 1900–2017 with at least one active reference	*WoS Core Collection* 2018 with at least one active reference	*Scopus* 2020 with ≥2 active references	*Scopus* 2020 with ≥2 references
Dataset size	45 million	Not specified (~3.6M)	3,034,904	3,121,740

Table 1. Main characteristics of the most recent document-level classification algorithms.

The main strength of this algorithm lies in its simplicity, which allows for rapid execution with limited resources. However, its restriction to single-category assignments limits its usefulness for research evaluation, as acknowledged by the author, although it shows potential for descriptive bibliometric and science-of-science studies. Precision loss is also admitted, as references that are not classifying are excluded, and applicability is limited to papers with active references. Notably, the evaluation of accuracy included small collections of articles manually categorized by their authors.

The Milojević proposal restriction to single-category assignments limits its usefulness for research evaluation

6. Multi-Generation Parametric Model

This proposal, described in *"Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research"* (**Glänzel** *et al.*, 2021), hereafter referred to as the Glänzel proposal, is a document-level (re)classification algorithm that successfully integrates several of the approaches reviewed earlier while introducing new concepts.

[1] Unlike the other algorithms in the table, this approach consists of two parallel classifications—one into the 14 broad areas of the NSF WebCASPAR and another into the 252 *WoS* categories. It is not a single hierarchical classification in which the 14 broad areas are subdivided into 252 narrower categories.

It builds upon *WoS* journal classifications adapted to the modified Leuven-Budapest scheme (**Glänzel** *et al.*, 2016), which comprises 16 categories and 74 subfields, compatible with *WoS* and *JCR*. While the model is theoretically based on multi-generation citation relationships, in practice it applies only two generations (direct references and their references). Three weighting models were tested: M1 (first generation only), M2 (second generation only), and M3 (weighted combination: 0.618 for first-generation references and 0.382[2] for second-generation references).

Unlike Milojević's approach, this model allows up to three weighted and normalized category assignments, with the sum of weights equal to 1. A threshold rule is applied: a new category is not assigned if its weight is less than two-thirds of the previous one. Assignments to the *WoS* "Multidisciplinary Sciences" category (X0) are explicitly excluded.

> The Glänzel model allows up to three weighted and normalized category assignments, with the sum of weights equal to 1, but a threshold rule is applied: a new category is not assigned if its weight is less than two-thirds of the previous one

Precision analyses determined that the M3 model yielded the best results. Its limitations include the inability to reclassify documents with few or no references, and difficulties in accurately reclassifying highly multidisciplinary articles, for which full-text analysis is suggested as a complementary approach.

7. *M3-AWC-0.8* Algorithm

Presented in *"New fractional classifications of papers based on two generations of references and on the ASJC Scopus scheme"* (**Álvarez-Llorente** *et al.*, 2024), this proposal builds on Glänzel's M1–M3 models and seeks to improve them by testing various adjustments.

First, a "smoothing" parameter is introduced to correct for imbalances between the number of first- and second-generation references, mitigating thematic bias caused by disciplinary citation practices. Three thresholds (½, ⅔, ⅘) were tested for allowing multiple category assignments, replacing Glänzel's fixed ⅔ rule. In addition, two counting methods were compared: full counting (assigning a weight of 1 to each category of a multi-assigned journal) and fractional counting (assigning a weight of 1/N when a journal belongs to N categories).

This algorithm uses the Fractional *ASJC* scheme (**Álvarez-Llorente** *et al.*, 2023) —the *Scopus ASJC* scheme with the multidisciplinary area and miscellaneous categories removed— and allows up to five simultaneous assignments.

[2] These values follow the criteria P1 + P2 = 1 and P12 = P2, whereby the second generation is assigned a reduced weight equal to the square of the first.

Extensive analysis of these variations focused especially on publications in multidisciplinary and miscellaneous journals, using comparisons with the *AAC* (*Author's Assignation Collection*) as a benchmark. Results showed that the most effective configuration was the M3 two-generation model with smoothing, fractional counting, and the 0.8 threshold, which gives the algorithm its name.

The study concludes that *M3-AWC-0.8* classifications are more homogeneous than journal-based classifications (adapted to Fractional *ASJC*), more consistent with *AAC*, and exhibit other desirable scientometric properties

The study concludes that *M3-AWC-0.8* classifications are more homogeneous than journal-based classifications (adapted to Fractional *ASJC*), more consistent with *AAC*, and exhibit other desirable scientometric properties. Its main limitation, once again, lies in the significant number of publications with very few references; documents with fewer than three active references across both generations were deliberately excluded from reclassification.

8. *U1-F-0.8* Algorithm

The most recent proposal, described in *"New paper-by-paper classification for Scopus based on references reclassified by the origin of the papers citing them"* (**Álvarez-Llorente** *et al.*, 2025), hereafter referred to as the *U1-F-0.8* algorithm, is closely linked to **Álvarez-Llorente** *et al.* (2024). Like *M3-AWC-0.8*, it uses the Fractional *ASJC* scheme and allows up to five weighted category assignments per article, facilitating comparison between the two algorithms and with the manual *AAC* classification.

U1-F-0.8 is a one-generation reference-based classification algorithm but introduces a novel idea: citation relationships are considered bidirectional. While a citing article is influenced by the subject category of the cited article, the cited work is also thematically "attracted" toward the citer. To operationalize this, a pre-classification step assigns categories to references based on the citing articles. Unlike traditional methods, this enables the classification of both active and inactive references, thereby expanding the number of articles that can be reclassified. Nevertheless, a minimum of three references is again required.

After pre-classification, the algorithm performs iterative reclassification. In each iteration, weights from categories not originally associated with the journal are removed. This process is repeated six times[3], followed by a seventh iteration without restrictions, producing two final versions: Journal Limited (JL) and Unlimited (U1). Additional adjustments

Additional adjustments include three thresholds for multiple assignments (½, ⅔, ⅘) and the option of fractional normalization, where assignment weights are divided by the number of references in each document to balance disciplinary citation differences

[3] It is repeated six times for the test dataset used in the validation of the algorithm. In practice, however, the iteration continues until the algorithm converges, that is, when the difference between the generated classification and that obtained in the previous iteration becomes sufficiently small.

include three thresholds for multiple assignments (½, ⅔, ⅘) and the option of fractional normalization, where assignment weights are divided by the number of references in each document to balance disciplinary citation differences. The algorithm is described in **Álvarez-Llorente** *et al.* (2025) from a formal algorithmic perspective and with a somewhat more mathematical approach. Nevertheless, to facilitate its interpretation, Appendix 1 provides a simplified explanation.

Results showed that the most effective configuration was obtained in the seventh iteration with fractional normalization and the 0.8 threshold —hence *U1-F-0.8*—.

Compared to *M3-AWC-0.8*, the classifications produced are similar, but *U1-F-0.8* offers higher consistency with *AAC* and other scientometric benchmarks, along with the significant advantage of broader applicability to a larger set of publications. This positions it as a viable candidate for adoption by major bibliographic databases for article-level classification. A deeper scientometric analysis of this algorithm is provided by **Peña-Rocha** *et al.* (2025).

Compared to *M3-AWC-0.8*, the classifications produced are similar, but *U1-F-0.8* offers higher consistency with *AAC* and other scientometric benchmarks, along with the significant advantage of broader applicability to a larger set of publications

9. Conclusions

This study began by examining the challenges faced by major bibliographic databases such as *WoS* and *Scopus* when categorizing the articles they index, since they do so by extending journal-level classifications to individual publications.

Since the emergence of these databases, numerous algorithms for individual article classification have been proposed, each with different purposes. Automatic clustering and community detection algorithms, while useful for studying research fronts, present key drawbacks: they generate unstable classifications with a high degree of randomness, rely on schemes that diverge significantly from those used by the databases, and often require manual labeling. As a result, they are not suitable for establishing stable categorizations of publications.

As an alternative, algorithms that (re)classify articles within the same thematic schemes as the databases appear to be a viable option. To determine the thematic content of articles, approaches based on various types of citation relationships and text analysis have been proposed, although the latter are generally too complex to be applied to large-scale datasets. Among these, direct citation relationships emerge as the most practical solution.

This tendency is reflected in the most recent proposals we have reviewed. **Milojević**'s (2020) *Practical Method to Reclassify WoS Articles* is limited by its restriction to single-category assignments, which reduces precision. The *Multi-Generation Parametric Model* by **Glänzel** *et al.* (2021) overcomes this limitation and represents a promising approach, but it remains constrained by the large number of articles with few or no references, which cannot be reclassified accurately. Building on this, **Álvarez-Llorente** *et al.* (2024) introduced the *M3-AWC-0.8* algorithm, which improves normalization and yields more balanced classifications, though it still suffers from the same restriction.

Finally, **Álvarez-Llorente** *et al.* (2025) proposed the *U1-F-0.8* algorithm, which advances beyond this limitation by reinterpreting the directionality of citation relationships. It achieves higher consistency with the *AAC* and incorporates other desirable scientometric features, making it the most effective proposal to date for document-level classification. As such, it stands as the most promising alternative to journal-based classifications in major bibliographic databases.

10. References

Althouse, B. M.; **West, J. D.**; **Bergstrom, C.T.**; **Bergstrom, T.** (2009). Differences in impact factor across fields and over time. *Journal of the Association for Information Science and Technology, 60*(1), 27–34.
https://doi.org/10.1002/asi.20936

Álvarez-Llorente, J. M. (2025). Nuevos algoritmos de clasificación de documentos científicos individuales basados en referencias para mejorar los análisis cienciométricos en las grandes bases de datos de ciencia [Doctoral thesis, University of Extremadura]. Institutional Repository of the University of Extremadura.

Álvarez-Llorente, J. M.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2023). Creating a collection of publications categorized by their research guarantors into the Scopus ASJC scheme. *Profesional de la Información*, 32(7).
https://doi.org/10.3145/epi.2023.dic.04

Álvarez-Llorente, J. M.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2024). New fractional classifications of papers based on two generations of references and on the ASJC Scopus scheme. *Scientometrics, 129*(6), 3493–3515.
https://doi.org/10.1007/s11192-024-05030-2

Álvarez-Llorente, J. M.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2025). New paper-by-paper classification for Scopus based on references reclassified by the origin of the papers citing them. *Journal of Informetrics*, *19*(2), 101647.
https://doi.org/10.1016/j.joi.2025.101647

Andersen, J. P. (2023). Field-level differences in paper and author characteristics across all fields of science in Web of Science, 2000-2020. *Quantitative Science Studies, 4*(2), 394–422.
https://doi.org/10.1162/qss_a_00246

Blondel, V. D.; **Guillaume, J. L.**; **Lambiotte, R.**; **Lefebvre, E.** (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, *2008*(10), P10008.
https://doi.org/10.1088/1742-5468/2008/10/P10008

Bornmann, L.; **Leydesdorff, L.** (2017). Skewness of citation impact data and covariates of citation distributions: A large-scale empirical analysis based on Web of Science data. *Journal of Informetrics, 11*(1), 164-175.
https://doi.org/10.1016/j.joi.2016.12.001

Bornmann, L.; **Tekles, A.**; **Leydesdorff, L.** (2019). How well does I3 perform for impact measurement compared to other bibliometric indicators? The convergent validity of several (field-normalized) indicators. *Scientometrics, 119*(2), 1187-1205.
http://dx.doi.org/10.1007/s11192-019-03071-6

Boyack, K. W.; **Klavans, R.** (2010). Co-citation analysis, bibliographic coupling, and direct citation: Which citation approach represents the research front most accurately? *Journal of the Association for Information Science and Technology, 61*(12), 2389–2404.
https://doi.org/10.1002/asi.21419

Boyack, K. W.; **Klavans, R.** (2020). A comparison of large-scale science models based on textual, direct citation and hybrid relatedness. *Quantitative Science Studies (1)4,* 1570–1585.
https://doi.org/10.1162/qss_a_00085

Boyack, K. W.; **Newman, D.**; **Duhon, R. J.**; **Klavans, R.**; **Patek, M.**; **Biberstine, J. R.**; **Schijvenaars, B.**; **Skupin, A.**; **Ma, N.**; **Börner, K.** (2011). Clustering more than two million biomedical publications: Comparing the accuracies of nine text-based similarity approaches. *PLoS One, 6*(3), e18029.
https://doi.org/10.1371/journal.pone.0018029

Boyack, K. W.; **Small, H.**; **Klavans, R.** (2013). Improving the Accuracy of Co-citation Clustering Using Full Text. *J Am Soc Inf Sci Tec, 64*: 1759–1767.
https://doi.org/10.1002/asi.22896

Chumachenko, A.; **Kreminskyi, B.**; **Mosenkis, I.**; **Yakimenko, A.** (2022). Dynamical entropic analysis of scientific concepts. *Journal of Information Science, 48*(4), 561–569.
https://doi.org/10.1177/0165551520972034

Clauset, A.; **Newman, M.**; **Moore, C.** (2004). Finding community structure in very large networks. *Physical Review E, 70*(6).
https://doi.org/10.1103/physreve.70.066111

De-Moya-Anegón, F.; **Herrero-Solana, V.**; **Jiménez-Contreras, E.** (2006). A connectionist and multivariate approach to science maps: the SOM, clustering and MDS applied to library and information science research. *Journal of Information Science, 32*(1), 63–77.
https://doi.org/10.1177/0165551506059226

Ding, J.; **Ahlgren, P.**; **Yang, L.**; **Yue, T.** (2018). Disciplinary structures in Nature, Science and PNAS: Journal and country levels. *Scientometrics*, 116(3), 1817–1852.
https://link.springer.com/article/10.1007/s11192-018-2812-9

Eykens, J.; **Guns, R.**; **Engels, T. C. E.** (2019). Article level classification of publications in sociology: An experimental assessment of supervised machine learning approaches. In: *Proceedings of the 17th International Conference on Scientometrics & Informetrics*, Rome (Italy), 2–5 September, 738–743.
https://hdl.handle.net/10067/1630240151162165141

Fang, H. (2015). Classifying Research Articles in Multidisciplinary Sciences Journals into Subject Categories. *Knowledge Organization, 42*(3), 139–153.
https://doi.org/10.5771/0943-7444-2015-3-139
Glänzel, W.; **Schubert, A.**; **Czerwon, H.** (1999a). An item-by-item subject classification of papers published in multidisciplinary and general journals using reference analysis. *Scientometrics, 44*(3), 427–439.
https://doi.org/10.1007/bf02458488

Glänzel, W.; **Schubert, A.**; **Schoepflin, U.**; **Czerwon, H.** (1999b). An item-by-item subject classification of papers published in journals covered by the SSCI database using reference analysis. *Scientometrics, 46*(3), 431–441.
https://doi.org/10.1007/BF02459602

Glänzel, W.; **Thijs, B.**; **Chi, P. S.** (2016). The challenges to expand bibliometric studies from periodical literature to monographic literature with a new data source: the book citation index. *Scientometrics, 109*, 2165–2179.
https://doi.org/10.1007/s11192-016-2046-7

Glänzel, W.; **Thijs, B.**; **Huang, Y.** (2021). Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research. In: W. Glänzel, S. Heeffer, P. S. Chi, R. Rousseau, *Proceedings of the 18th International Conference of the International Society of Scientometrics and Informetrics (ISSI 2021*), Leuven University Press, 453–464.
https://kuleuven.limo.libis.be/discovery/fulldisplay?docid=lirias3394551&context=SearchWebhook&vid=32KUL_KUL:Lirias&search_scope=lirias_profile&tab=LIRIAS&adaptor=SearchWebhook&lang=en

Gläser, J.; **Glänzel, W.**; **Scharnhorst, A.** (2017). Same data—Different results? Towards a comparative approach to the identification of thematic structures in science. *Scientometrics, 111*(2), 981–998.
https://doi.org/10.1007/s11192-017-2296-z

Glenisson, P.; **Glänzel, W.**; **Janssens, F.**; **De-Moor, B.** (2005). Combining full text and bibliometric information in mapping scientific disciplines. *Information Processing & Management, 41*(6), 1548–1572.
https://doi.org/10.1016/j.ipm.2005.03.021

Gómez-Crisóstomo, M. R. (2011). *Study and comparison of the Web of Science and Scopus (1996-2007)* [*Doctoral thesis, University of Extremadura*]. Institutional Repository of the University of Extremadura.

Guerrero-Bote, V. P.; **De-Moya-Anegón, F.** (2012). A further step forward in measuring journals' scientific prestige: The SJR2 indicator. *Journal of informetrics*, *6*(4), 674-688. *https://doi.org/10.1016/j.joi.2012.07.001*

Guerrero-Bote, V.P.; **De-Moya-Anegón, F.**; **Herrero-Solana, V.** (2002). Document organization using Kohonen's algorithm. Information Processing and Management, 38(1), pp. 79-89.
https://doi.org/10.1016/S0306-4573(00)00066-2

Guerrero-Bote, V. P.; **Zapico-Alonso, F.**; **Espinosa-Calvo, M. E.**; **Gómez-Crisóstomo, R.**; **De-Moya-Anegón, F.** (2007). Import-export of knowledge between scientific subject categories: The iceberg hypothesis. *Scientometrics, 71*(3), 423–441.
https://doi.org/10.1007/s11192-007-1682-3

Hassan-Montero, Y.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2022). SCImago Graphica: a new tool for exploring and visually communicating data. *Profesional de la información, 31*(5), e310502. *https://doi.org/10.3145/epi.2022.sep.02*

Hassan-Montero, Y.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2014). Graphical interface of the SCImago Journal and Country Rank: an interactive approach to accessing bibliometric information. *El profesional de la información*, 23(3). *https://doi.org/10.3145/epi.2014.may.07*

Huang, Y.; **Glänzel, W.**; **Thijs, B.**; **Porter, A. L.**; **Zhang, L.** (2021). *The comparison of various similarity measurement approaches on interdisciplinary indicators* (pp. 1–24). FEB - KU Leuven

Janssens, F.; **Glänzel, W.**; **De-Moor, B.** (2008). A hybrid mapping of information science. *Scientometrics, 75*(3), 607–631.
https://doi.org/10.1007/s11192-007-2002-7

Janssens, F.; **Leta, J.**; **Glänzel, W.**; **De-Moor, B.** (2006). Towards mapping library and information science. *Information Processing & Management, 42*(6), 1614–1642. *https://doi.org/10.1016/j.ipm.2006.03.025*

Janssens, F.; **Zhang, L.**; **De-Moor, B.**; **Glänzel, W.** (2009). Hybrid clustering for validation and improvement of subject-classification schemes. *Information Processing & Management, 45*(6), 683–702.
https://doi.org/10.1016/j.ipm.2009.06.003

Javitz, H.; **Grimes, T.**; **Hill, D.**; **Rapoport, A.**; **Bell, R.**; **Fecso, R.**; **Lehming, R.** (2010). *U.S. Academic Scientific Publishing. Working paper SRS 11-201*. Arlington, VA: National Science Foundation, Division of Science Resources Statistics.

Kandimalla, B.; **Rohatgi, S.**; **Wu, J.**; **Giles, C. L.** (2021). Large scale subject category classification of scholarly papers with deep attentive neural networks. *Frontiers in Research Metrics and Analytics, 5*, 600382.
https://doi.org/10.3389/frma.2020.600382

Klavans, R.; **Boyack, K. W.** (2005). Identifying a better measure of relatedness for mapping science. *Journal of the Association for Information Science and Technology, 57*(2), 251-263.
https://doi.org/10.1002/asi.20274

Klavans, R.; **Boyack, K. W.** (2006). Quantitative evaluation of large maps of science. *Scientometrics, 68*, 475–499.
https://doi.org/10.1007/s11192-006-0125-x
Klavans, R.; **Boyack, K. W.** (2016). Which Type of Citation Analysis Generates the Most Accurate Taxonomy of Scientific and Technical Knowledge? *Journal of the Association for Information Science and Technology, 68*(4), 984–998.
https://doi.org/10.1002/asi.23734

Lai, K.; **Wu, S.** (2005). Using the patent co-citation approach to establish a new patent classification system. *Information Processing & Management, 41*(2), 313–330.
https://doi.org/10.1016/j.ipm.2003.11.004

Lancho-Barrantes, B. S.; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2010a). The iceberg hypothesis revisited. *Scientometrics, 85*(2), 443–461.
https://doi.org/10.1007/s11192-010-0209-5

Lancho-Barrantes, B. S.; **Guerrero-Bote, V. P.**; **De-Moya Anegón, F.** (2010b). What lies behind the averages and significance of citation indicators in different disciplines? *Journal of Information Science, 36*(3), 371-382.
https://doi.org/10.1177/0165551510366077

Leydesdorff, L.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2010). Journal maps on the basis of Scopus data: A comparison with the Journal Citation Reports of the ISI. *Journal of the American Society for Information Science and Technology*, 61(2), 352-369.
https://doi.org/10.1002/asi.21250

Leydesdorff, L.; **De-Moya-Anegón, F.**; **Guerrero-Bote, V. P.** (2015). Journal maps, interactive overlays, and the measurement of interdisciplinarity on the basis of scopus data (1996–2012). *Journal of the Association for Information Science and Technology*, 66(5), 1001-1016.
https://doi.org/10.1002/asi.23243

Li, K.; **Chen, P.-Y.**; **Fang, Z.** (2019). Disciplinarity of software papers: A preliminary analysis. *Proceedings of the Association for Information Science and Technology (56*), 706–708.
https://doi.org/10.1002/pra2.143

Marshakova-Shaikevich, I. (2005). Bibliometric maps of field of science. *Information Processing & Management, 41*(6), 1534–1547.
https://doi.org/10.1016/j.ipm.2005.03.027

McGillivray, B.; **Astell, M.** (2019). The relationship between usage and citations in an open access mega-journal. *Scientometrics, 121*, 817–838.
https://doi.org/10.1007/s11192-019-03228-3

Milojević, S. (2020). Practical method to reclassify Web of Science articles into unique subject categories and broad disciplines. *Quantitative science studies, 1*(1), 183-206.
https://doi.org/10.1162/qss_a_00014

Opthof, T.; **Leydesdorff, L.** (2010). Caveats for the journal and field normalizations in the CWTS ("Leiden") evaluations of research performance. *Journal of informetrics*, 4(3), 423-430.
https://doi.org/10.1016/j.joi.2010.02.003

Peña-Rocha, M.; **Gómez-Crisóstomo, R.**; **Guerrero-Bote, V. P.**; **De-Moya-Anegón, F.** (2025). Bibliometrics effects of a new paper level classification. *Frontiers in Research Metrics and Analytics, 10*.
https://doi.org/10.3389/frma.2025.1531758

Rees-Potter, L. K. (1989). Dynamic thesaural systems: A bibliometric study of terminological and conceptual change in sociology and economics with application to the design of dynamic thesaural systems. *Information Processing & Management, 25*(6), 677–689.
https://doi.org/10.1016/0306-4573(89)90101-5

Sachini, E.; **Sioumalas-Christodoulou, K.**; **Christopoulos, S.**; **Karampekios, N.** (2022) AI for AI: Using AI methods for classifying AI science documents. *Quantitative Science Studies, 3*(4), 1119–1132.
https://doi.org/10.1162/qss_a_00223

Schildt, H.; **Mattsson, J.** (2006). A dense network sub-grouping algorithm for co-citation analysis and its implementation in the software tool Sitkis. *Scientometrics, 67*, 143–163.
https://doi.org/10.1007/s11192-006-0054-8

Shu, F.; **Julien, C.**; **Zhang, L.**; **Qiu, J.**; **Zhang, J.**; **Larivière, V.** (2019). Comparing journal and paper level classifications of science. *Journal of Informetrics, 13*(1), 202–225. *https://doi.org/10.1016/j.joi.2018.12.005*

Šubelj, L.; **Van Eck, N. J.**; **Waltman, L.** (2016). Clustering Scientific Publications Based on Citation Relations: A Systematic Comparison of Different Methods. *PLoS one, 11*(4), e0154404.
https://doi.org/10.1371/journal.pone.0154404

Thelwall, M.; **Pinfield, S.** (2024). The accuracy of field classifications for journals in Scopus. *Scientometrics, 129*(2), 1097–1117.
https://doi.org/10.1007/s11192-023-04901-4

Thijs, B.; **Huang, Y.**; **Glänzel, W.** (2021). *Comparing different implementations of similarity for disparity and variety measures in studies on interdisciplinarity*. FEB Research Report MSI_2103, Report No. MSI_2103.
https://lirias.kuleuven.be/retrieve/610314

Van Eck, N.J.; **Waltman, L.** (2010). Software survey: VOSviewer, acomputer program for bibliometric mapping. *Scientometrics, 84*(2), 523–538.
https://doi.org/10.1007/s11192-009-0146-3

Waltman, L.; **Boyack, K. W.**; **Colavizza, G.**; **Van Eck, N. J.** (2020). A principled methodology for comparing relatedness measures for clustering publications. *Quantitative Science Studies, 1*(2), 691-713.
https://doi.org/10.1162/qss_a_00035

Waltman, L.; **Van Eck, N. J.** (2012). A new methodology for constructing a publication-level classification system of science. *Journal of the Association for Information Science and Technology, 63*(12), 2378–2392.
https://doi.org/10.1002/asi.22748

Wang, Q.; **Waltman, L.** (2016). Large-scale analysis of the accuracy of the journal classification systems of Web of Science and Scopus. *Journal of Informetrics, 10*(2), 347-364.
https://doi.org/10.1016/j.joi.2016.02.003

Zhang, J.; **Shen Z.** (2024). Analyzing journal category assignment using a paper-level classification system: multidisciplinary sciences journals. *Scientometrics*, *129*, pp. 5963-5978.
https://doi.org/10.1007/s11192-023-04913-0

Zhang, L.; **Janssens, F.**; **Liang, L.**; **Glänzel W.** (2010). Journal cross-citation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics, 82*, 687–706.
https://doi.org/10.1007/s11192-010-0180-1

Zhang, L.; **Rousseau, R.**; **Glänzel, W.** (2016). Diversity of references as an indicator of the interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology, 67*(5), 1257-1265.
https://doi.org/10.1002/asi.23487

Zhang, L.; **Sun, B.**; **Shu, F.**; **Huang, Y.** (2022). Comparing paper level classifications across different methods and systems: an investigation of Nature publications. *Scientometrics, 127*(12), 7633–7651.
https://doi.org/10.1007/s11192-022-04352-3

Appendix 1. Algorithm *U1-F-0.8*