# SciNoBo: A Hierarchical Multi-Label Classifier of Scientific Publications

Nikolaos Gialitsis*
Athena Research and Innovation Center
Institute for Language and Speech Processing
Athens, Greece
ngialitsis@athenarc.gr

Sotiris Kotitsas†
Athena Research and Innovation Center
Institute for Language and Speech Processing
Athens, Greece
sotiris.kotitsas@athenarc.gr

Haris Papageorgiou‡
Athena Research and Innovation Center
Institute for Language and Speech Processing
Athens, Greece
haris@athenarc.gr

## ABSTRACT

Classifying scientific publications according to Field-of-Science (FoS) taxonomies is of crucial importance, allowing funders, publishers, scholars, companies and other stakeholders to organize scientific literature more effectively. Most existing works address classification either at venue level or solely based on the textual content of a research publication. We present SciNoBo, a novel classification system of publications to predefined FoS taxonomies, leveraging the structural properties of a publication and its citations and references organized in a multilayer network. In contrast to other works, our system supports assignments of publications to multiple fields by considering their multidisciplinarity potential. By unifying publications and venues under a common multilayer network structure made up of citing and publishing relationships, classifications at the venue-level can be augmented with publication-level classifications. We evaluate SciNoBo on a dataset of publications extracted from Microsoft Academic Graph, and we perform a comparative analysis against a state-of-the-art neural-network baseline. The results reveal that our proposed system is capable of producing high-quality classifications of publications.

## CCS CONCEPTS

**• Applied computing → Document metadata**; **Digital libraries and archives**; **• Computing methodologies → Topic modeling**; **• Mathematics of computing → Graph algorithms**.

## KEYWORDS

field of science publication classification, multilayer network, label propagation, scholarly data, digital libraries, neural networks, hierarchical classification, multi-label classification

**ACM Reference Format:**
Nikolaos Gialitsis, Sotiris Kotitsas, and Haris Papageorgiou. 2022. SciNoBo: A Hierarchical Multi-Label Classifier of Scientific Publications. In *Companion Proceedings of the Web Conference 2022 (WWW '22 Companion), April 25–29, 2022, Virtual Event, Lyon, France.* ACM, New York, NY, USA, 10 pages. https://doi.org/10.1145/3487553.3524677

## 1 INTRODUCTION

It is estimated that the overall volume of scientific publications doubles every 17.3 years, and every 12.9 years in physical and technical sciences [6]. In order to manage the wealth and growth rate of knowledge, multiple literature databases have emerged covering different perspectives: Microsoft Academic [48], Scopus [3], Web Of Science (WoS) [4], SemanticScholar, Crossref [18], OpenCitations [32], OpenAIRE [28], Dimensions [19], ScienceDirect [1] as well as specialized databases such as PubMed [2] and the Computer Science Ontology (CSO) [37].

To empower search engines (such as Google Scholar, Semantic Scholar, ArnetMiner [42], InCites, Pub Finder [45]) recommendation systems [8], science and innovation monitoring efforts [5, 15, 25, 41, 46], and to normalize bibliometric indices [9, 10, 34, 44], literature databases often adopt classification schemes for tagging entities (venues/publications/projects) in respect to their thematic contents: scientific areas, subject fields or topics. Entities can either be classified to one or more scientific fields from a predefined list, or can be assigned a list of representative keywords by the author(s) or/and the journal editor(s).

Scientific fields are often organized hierarchically into a taxonomy in which the top-levels represent broad subject areas and disciplines such as *social sciences* and *engineering and technology* and the bottom levels represent fine-grained and specialized subfields such as *optics & laser technology*. Examples of taxonomies for scientific field classification are the: All Science Journal Classification (ASJC) System, International Classification of Diseases (ICD) [22], Frascati Manual Classification [29], Medical Subject

[*] designed and developed the initial implementation and the theoretical formalism of SciNoBo, undertook the literature review and the documentation of the proposed method.
[†] worked out almost all technical details, implemented the baseline method dann, ran the experiments (including data collection and pre-processing) and comparative analysis, documented and reviewed the results.
[‡] conceived the study and supervised the project.

*WWW '22 Companion, April 25–29, 2022, Virtual Event, Lyon, France*
© 2022 Association for Computing Machinery.
ACM ISBN 978-1-4503-9130-6/22/04…$15.00
https://doi.org/10.1145/3487553.3524677

[1] ScienceDirect [Internet]. Elsevier [1997] - [cited June 6, 2022] Available from: https://www.sciencedirect.com/
[2] PubMed [Internet]. Bethesda (MD): National Library of Medicine (US). [1946] - [cited June 6, 2022]. Available from: https://www.ncbi.nlm.nih.gov/pubmed/

Headings (MeSH) [3], WoS Categories and Subject Areas [4], Scopus Subject Areas [5], European Science Vocabulary (EuroSciVoc) [6] and Microsoft Academic Graph Concepts [38].

Recently, the scientometrics community has been shifting their focus from venue-level (journals/conferences) to publication-level classification systems, as evidenced by a growing line of research [12, 20, 21, 33, 35, 36, 38, 47]. Comparative studies [31, 39] have shown that classification systems at the publication-level are more precise than venue-level counterparts as they naturally provide a higher degree of granularity, which can prove advantageous in certain applications.

Despite the fact that publication-level systems have been proposed several decades ago [13, 14], venue-level classifications, including WoS and Scopus [2, 26, 30], are still in use, since they are more easily curated by publishers and editors due to their smaller volume and more static nature. An extensive summary of approaches for venue-level classification and their limitations has been presented by Archambault et al. [2].

The majority of methods proposed for automatic FoS categorization perform some form of clustering on publications in order to map science or to identify topics from scratch. These include, topic modeling approaches [17, 42] and the grouping of publications via citation networks [40, 43] or bibliographic coupling [47]. Even though such unsupervised approaches are useful in many cases, they solve a different task to the one we focus on in this publication, which is the classification of publications according to a predefined FoS taxonomy

Our method SciNoBo, contributes to the domain of taxonomy-driven FoS classification in more than one way. First and foremost, SciNoBo classifies publications across all disciplines, in contrast to other works that focus on a specific domain. Secondly, it is suitable for handling multidisciplinarity as it can be applied in both multiclass and multi-label classification settings. Furthermore, SciNoBo supports assignments to multiple levels of detail within a given FoS taxonomy by encoding hierarchical relationships among FoS labels. Moreover, SciNoBo classifies publications by requiring minimal metadata. A publication can be classified from the first day it becomes available online, and later as more metadata gradually become available, SciNoBo can classify the publication again by taking into account richer relationships. Lastly, we employ a new FoS taxonomy that extends OECD disciplines with sciencemetrix FoS codes.

The publication is structured as follows: We start by reviewing different approaches found in the literature for classifying publications according to a predefined FoS taxonomy. In section 3, we discuss the main limitations of existing approaches and the drivers for additional research. We then proceed in section 4 and describe in-detail the proposed methodology and its mathematical formulation. In the next section, we report on conducted experiments and implementation details. Next, we present results and discuss how SciNoBo performs against a neural state-of-the-art baseline. Finally, we reach conclusions and propose directions for subsequent research and improvements.

# 2 RELATED WORK

Approaches for publication-level FoS classification mostly rely on metadata including titles, author-keywords and abstracts, since full text is often unavailable or locked behind a paywall. We distinguish two main approaches found in literature: keyword extraction methods, and machine learning methods.

## 2.1 Keyword extraction methods

Keyword extraction methods have been applied with the goal of identifying small sets of representative words, phrases, or n-grams to associate with a predefined set of FoS. According to the similarity scores (e.g. measuring overlap, or some vector space similarity such as the angle between word-vectors), the FoS candidates are ranked and the publication is classified to the best matching field(s).

For example, Salatino et al. [35] proposed a text-based classifier for classifying publications to one or more research area(s) from the Computer Science Ontology (CSO [37]). N-grams extracted from each abstract are matched to FoS labels by means of Levenshtein similarity as well as by the cosine similarity between their pre-trained word2vec embeddings. Even though their approach performs multi-label classification and supports hierarchical assignments through CSO child-parent relationships, intensive post-processing effort is needed to detect and filter-out false-positives. Also, the classifications are confined to the Computer Science domain.

Shen et al. [38] describe the concept tagging of publications within Microsoft Academic Graph (MAG). Publications are represented based on both graph structural and textual information by leveraging metadata including venue names, titles, keywords and abstracts, in addition to the metadata of their neighbors within the graph. Similarly, MAG concepts are represented by the first paragraph of their corresponding Wikipedia entry (concepts are derived from Wikipedia). The vector embeddings of both representations are compared through cosine similarity and the publication is classified to the concept (FoS) if the similarity exceeds a predefined threshold. Nevertheless, the representation of a MAG concept directly depends on its Wikipedia entry, whereas for most taxonomies, only word labels represent the classes. Furthermore, empirical weights and heuristics are applied therein which prohibits complete reusability.

## 2.2 Machine learning methods

### 2.2.1 Traditional methods.

This category encompasses supervised machine learning methods i.e. labeled examples are required in order to train the model. These were some of the first approaches towards the automatic classification of publications according to a pre-defined FoS taxonomy.

Caragea et al. [7] classify publications to one out of six FoS categories from the CiteSeer literature database by taking advantage of citation relationships. The citation contexts of publications (citing and cited) are used in order to train different variations of the Multinomial Naive Bayes classifier. However, this approach does not support multi-label assignments and is only tested on a classification scheme containing few non-hierarchical FoS categories.

Domain-specific FoS classification systems have also been sporadically developed. Lukasik et al. [27] examine a combination of

[3] MeSH
[4] WoS
[5] Scopus Subject Areas
[6] EuroSciVoc

Naive Bayes and kNN algorithm on the hierarchical multi-label classification of publications according to Mathematical Subject Headings (MSC). Similarly, Kurach et al.[24] construct classifier ensembles in order to assign MeSH terms to biomedical publications from the Pubmed Central database. They evaluate various combinations of machine learning methods in a supervised multi-label classification setting. However, their study does not account for hierarchical relationships among FoS labels.

### 2.2.2 *Deep learning methods.*

Nowadays, traditional machine learning approaches for FoS classification have mostly been superseded by deep learning methods.

Kandimalla et al. [21] employ a deep neural encoder equipped with the attention mechanism in order to classify publications to WoS subject categories. They keep the most representative words per abstract in terms of TF-IDF score and subsequently map them to pre-trained word embeddings before feeding them to the neural network. However their approach does not take into account the hierarchical relationships of the FoS taxonomy and finds difficulty in discriminating between FoS categories with similar vocabulary.

In [33], a modified character-level Convolutional Neural Network (CNN) is compared against the more traditional approaches of bibliographic coupling and direct citation for the problem of FoS classification. They evaluated the performance of each method separately on a publication-level human-curated test dataset comprising of 200 publications in total, half randomly sampled from disciplinary journals, and the other half from multidisciplinary. For training the CNN, the training labels used were direct extensions of the venue label. No clear winning method was announced in the comparative study as all resulted in similar performance. Multiple sources of metadata were given as input to the CNN, including authors' affiliations, referenced venue names, reference titles, abstract, keywords, title, and subfield classification of references. Most of these however are rarely available. Furthermore, the text-based input is concatenated/truncated arbitrarily and does not take into account term semantics.

Hoppe et al.[20] combine a BiLSTM neural network architecture with Knowledge Graphs (KG) in order to represent class labels by their KG-embeddings. The intuition is that by including the KG embeddings of the classes, in addition to their word labels, the FoS classifier will achieve higher performance. They demonstrate the above, on the multi-label classification of $\sim 92K$ ArXiv publications represented by DBpedia embeddings. Nevertheless, the relationship between the FoS labels is not taken into account as the multi-label classification problem is decomposed into independent binary classifications.

In the position paper HierClasSArt [1], knowledge graphs (KGs) were proposed, in addition to neural networks, for the FoS classification of publications to mathematical topics from the zbMATH database. KG-embeddings leveraging both textual and structural metadata can be derived from the available metadata, and successively be provided as input to a deep neural network. Hierarchical relations among entities may be captured through the help of NLP methods or human experts, but the exact details are not thoroughly described. While such a methodology can potentially be transferred to other domains, only the domain of mathematics was explored in the study.

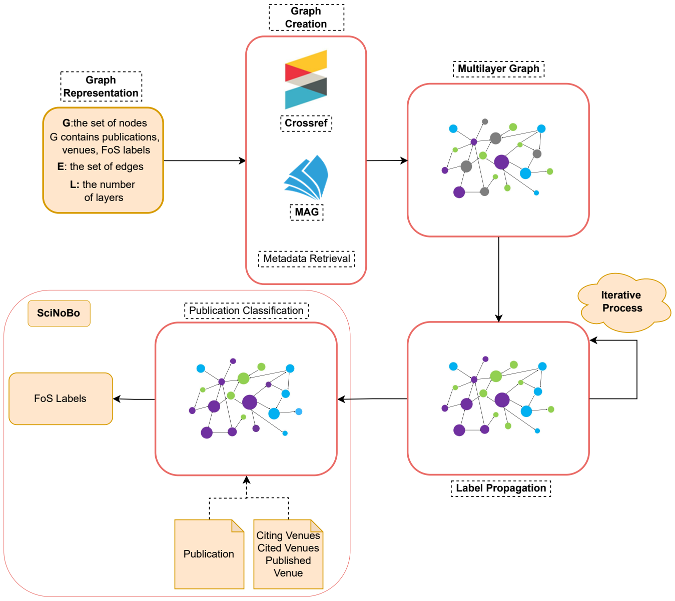

**Figure 1: Illustration of the proposed method. After we define our graph in Graph Representation, we retrieve metadata and construct it. The result is a Multilayer Graph and after Label Propagation, we can input a publication along with the required metadata to retrieve the FoS labels.**

## 3 MOTIVATION

Existing FoS classification approaches completely ignore, or face significant difficulties when dealing with multidisciplinarity, either at the venue or at the publication-level. Moreover, nearly all of them depend on textual content which, when available, is prone to concept drift, discourse norms in specific fields and multilinguality specificities. Moreover, several approaches confine classifications to a specific discipline or lack generalization capabilities. In addition, hierarchical relationships between FoS labels are often under-utilized or ignored.

Therefore, there is still a need for developing systems for efficient multi-label classification at the publication level. Our motivation is that by taking into account both the citing/publishing relationships at the publication-level as well as at the venue-level, we will be able to provide "context-aware" classifications without considering publication content. In contrast to other methods, we propose a multiclass multi-label classification approach assigning research publications to one or more FoS codes capturing the increasing multidisciplinarity in literature.

## 4 SCINOBO

The method we propose is based on the assumption that a publication [7] mostly cites thematically related publications. We bridge venues (journals/conferences) and publications by constructing a multilayer network (graph) in which venues are represented by nodes, and venue-venue edges reflect citing-cited relationships in their respective publications. SciNoBo classifies publications to

[7] We use the term "publication" to refer to all peer-review research works published in journals or conferences.

one or more FoS based on the publishing venues of the publications it references (out-citations) and the publishing venues of the publications it gets cited by (in-citations). Therefore, SciNoBo classifies publications with minimal metadata utilizing only journal or conference names as well as citing information. Figure 1 illustrates the steps followed to create SciNoBo. Each step of the approach is analytically covered in the following subsections.

## 4.1 Graph Representation

SciNoBo unifies multiple types of relationships (edges) between entities as well as multiple types of entities under a common framework of operations represented as a multilayer network [8] (see Figure 2). We consider a multilayer network $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathcal{L})$ where $\mathcal{V}$ contains the set of publications $\mathcal{P}$, the set of venues $\mathcal{J}$, and the set of scientific fields $\mathcal{F}$. Or equivalently: $\mathcal{V} = \mathcal{P} \bigcup \mathcal{J} \bigcup \mathcal{F}$. The symbol $\mathcal{E}$ represents the set of edges (links) between nodes; and $\mathcal{L}$ is the set of layers capturing different types of relationships between nodes. Since the network has multiple layers, each edge belongs to one layer $\ell \in \mathcal{L}$ and has a weight $w \in \mathbb{R}^+$. We can represent all edges in the network using 4-tuples as: $\mathcal{E} = \{(u, v, \ell, w); u, v \in \mathcal{V}, \ell \in \mathcal{L}, w \in \mathbb{R}^+\}$ and $(u, v, \ell, w) \neq (v, u, \ell, w)$ (directed edges). Edges in layer $\ell \in \mathcal{L}$ represent a particular type of connection among nodes, and two nodes $u, v$ might be connected by edges in multiple layers.

We formulate the task of scientific field classification as a link-prediction problem in the multilayer network. The goal is to predict edges between publication-nodes and scientific-field nodes: $\{(u, v, \ell_0, w); u \in \mathcal{P}, v \in \mathcal{F}, \ell_0 \in \mathcal{L}, w \in \mathbb{R}^+\}$.

An edge between two publications $(p_i, p_j)$ at $\ell_1$ means that publication $p_i$ cites publication $p_j$. An edge at layer $\ell_2$ connects a publication to its publishing venue(s). We also define edges between venues at $\ell_3$: $\{(u, v, \ell, w); u, v \in \mathcal{J}, \ell \in \mathcal{L}, w \in \mathbb{R}^+\}$; where $w$ is the number of publications which have been published in venue $u$ and cite (reference) publications published in venue $v$. The weight of an edge $(v, f)$ at $\ell_4$ corresponds to how thematically relevant the publications published in $v$ are to the scientific field $f$. Finally, $\ell_5$+ layers represent hierarchical relationships among labels.

## 4.2 Graph Creation

SciNoBo network was populated by exploiting CROSSREF [9] and Microsoft Academic Graph (MAG) [10]. CROSSREF contains more than 120 million publications and MAG contains approximately 250 million records. We retrieve all the publications that were published between $2016 - 2021$, along with their references [11] and their citations when available.

For every publication, the publishing venue is contained in the metadata. However this is not the case for the references and citations. As a result, for every publication we query its references and citations in CROSSREF/MAG (by taking the union of the metadata) and we retrieve the original metadata of the reference or the citation. Inherently, we can now create venue-venue relationships (edges at layer $\ell_3$) as in 4.1. The weight of a venue-venue edge is the amount of times a venue has referenced or being cited by another venue. An edge is created between two venues if they (i.e. their respective publications) cite each other more than 10 times [12]. Post graph creation, we normalize the weights of each node's outgoing edges to sum up to 1 by diving with the maximum weight of each neighborhood. The normalized weight of a venue-venue edge $(u, v)$ can be roughly interpreted as the probability of a publication published in $u$ to cite a publication published in $v$.

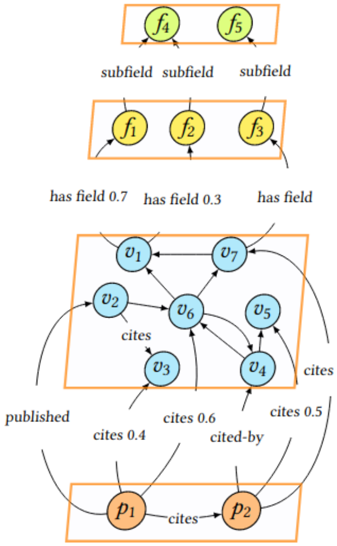

**Figure 2: Schematic representation of the multilayer network**

***Venue Deduplication.*** : A considerable challenge is dealing with naming inconsistencies in the reporting of venues in publication references/citations, or different instances of the same venue. This challenge is particularly prevalent in Crossref metadata since the published venue of each publication is being deposited by the authors. Our main goal is to create abbreviations for the names of the venues e.g. the "Empirical Methods in Natural Language Processing" conference should be mapped to EMNLP. Furthermore, different instances of venues should also be mapped to a unique venue abbreviation (e.g. EMNLP 2019, EMNLP 2020 etc. to EMNLP) [13]. In addition by performing an exploratory analysis on the names of the reported venues, we conclude that most of the abbreviations exist after the character '−' and inside parentheses.

While parsing the publications during the Graph Creation process, we keep a mapping from the full venue names to the venue

[8] Multilayer networks are data structures used to model complex interactions [23], ranging from Biomedicine [16], to Social Network Analysis [11].

[9] https://www.crossref.org/

[10] https://www.microsoft.com/en-us/research/project/microsoft-academic-graph/

[11] references retrieved were confined in a 10 year window. Older references should be omitted as it is customary for research papers to cite old works (e.g. in the introduction section) in order to provide general background knowledge whose themes may not directly coincide with the publication's.

[12] We observed that by using a lower weight threshold we introduce noise into the network by including low-signal venues of questionable quality. On the contrary, a weight threshold $t \sim 100$ leads into a large number of reputable conferences to being omitted as their individual instances (e.g. annual) are filtered before the venue deduplication and aggregation steps, discussed in (4.2), take place.

[13] The following preprocessing was applied to the names of the venues: Removal of latin, cardinal and ordinal numbers, dates, days, months, pre-specified words/phrases (e.g. "speech given", "oral talk" etc.), stopwords, special characters;adding a space when removing them and normalizing multiple spaces. The same preprocessing procedure is also applied during inference.

abbreviations we have identified, while creating the venue-venue edges. The mapping created is also used during inference time, to map the incoming venues to the abbreviated venue names in the multilayer graph of SciNoBo.

***Field-of-Science Taxonomy.*** : Our classification scheme is underpinned by the OECD disciplines/fields of research and development (FORD) classification scheme, developed in the framework of the Frascati Manual [14] and used to classify R&D units and resources in broad (first level(L1), one-digit) and narrower (second level(L2), two-digit) knowledge domains based primarily on the R&D subject matter. To facilitate a more fine grained analysis, we extend the OECD/FORD scheme by manually linking FoS labels of the sciencemetrix [15] classification scheme to OECD/FORD level-2 categories, creating a hierarchical 3-layer taxonomy. Table 1 provides statistics of our taxonomy.

**Table 1: Statistics of the extended OECD/FORD classification scheme.**

| Levels of FoS | Number of Labels |
|---|---|
| Level 1 | 6 |
| Level 2 | 42 |
| Level 3 | 174 |

sciencemetrix Classification also provides a list of Journal Classifications. We integrate this list, by mapping its journals to SciNoBo nodes and linking them with the relevant FoS codes. This mapping represents $\ell_4$ relationships, which are utilized to classify publications in FoS labels. Initially a small portion of venues have an FoS in Level-2 and Level-3. By utilizing Label Propagation, we aim to increase the venue label coverage.

## 4.3 Label Propagation

The intuition behind incorporating venues into the network, is that starting from a small set of seeds (venues with FoS labels), we can propagate the information to the rest of the network. The hypothesis is that a venue is more likely to express the FoS of its most referenced venues, an approach resembling a nearest-neighbor classification setting.

We assume that only a subset of venues $\mathcal{J}^* \subseteq \mathcal{J}$ has available FoS labels (i.e. venue-FoS edges in layer $\ell_4$). However, we do not consider these seed venue-FoS classifications to be ground-truth and we re-evaluate them dynamically during label propagation. By taking into account the network of venue-venue relationships, we enrich the initial FoS journal classifications described in 4.2 by inferring additional venue-FoS relationships. Consequently, previously single-labeled classifications may become multi-labeled after a few rounds of label propagation.

Label propagation is an iterative procedure. On each iteration, every venue-node aggregates the FoS labels of its neighbors in proportion to the venue-neighbor preference and neighbor-FoS preference. The preference of a venue towards one of its neighbors is expressed through the venue-venue normalized edge weight at $\ell_4$. A weight of 0 equals to no preference (i.e. publications do not cite publications published by the neighbor venue) whereas a weight of 1 equals to absolute preference (all publications cite publications published by the neighboring venue). These two preferences can be multiplied together, to estimate the expected weight between the venue, and each FoS node connected to its neighbors.

**Table 2: Complete graph statistics. Pre Label Propagation indicates the number of venues that have FoS labels per Level before Label Propagation. Post Label Propagation indicates the number of venues with FoS Labels per Level after we apply Label Propagation. Nodes represent the number of venue nodes in the multilayer graph.**

| Nodes | Edges | Pre Label Propagation<br>Lvl1-Lvl2-Lvl3 | | | Post Label Propagation<br>Lvl1-Lvl2-Lvl3 | | |
|---|---|---|---|---|---|---|---|
| 94752 | 3112953 | N/A | 84 | 14889 | 32049 | 32049 | 32324 |

Given layers:$\ell_t, \ell_{t+1}$the label propagation formula is:

$$\overset{\text{venue-FoS}}{w_{i,k}^{\ell_{t+1}}} = \sum_{j \in N_i^{\ell_t}} \sum_{k \in mathcal N_j^{\ell_{t+1}}} \overset{\text{venue-neighbor}}{w_{i,j}^{\ell_t}} \cdot \overset{\text{neighbor-FoS}}{w_{j,k}^{\ell_{t+1}}}$$

$$\text{for all citing venues in } \ell_t$$
$$\forall i \in \{u : \exists (u, \cdot, \ell_t, \cdot); u \in \mathcal{V}\}$$
$$\text{for each cited neighbor's FoS}$$
$$\forall k \in \{v : \exists (j, v, \ell_{t+1}, \cdot); j \in \mathcal{N}_i^{\ell_t}; v \in \mathcal{V}\}$$
$$\text{where } \mathcal{N}_i^{\ell_t} \text{ are venues cited by } i \in \ell_t$$

Complete statistics regarding the Graph (before and after Label Propagation) are presented in Table 2. We observe that Label Propagation achieves wide coverage between venue-FoS ($\ell_4$) edges.

## 4.4 Publication Classification

SciNoBo can assign FoS labels to individual publications through *citing/cited-by relationships and publishing/published-by relationships.* Publication-classification uses the same label propagation mechanism as the one presented in 4.3. Assume that each publication is represented by a unique node in the SciNoBo network. The goal is to connect each publication node to one or more FoS nodes in layer $\ell_0$. We have already discussed how venue-FoS relationships ($\ell_4$) can be established in the subsections 4.2 and 4.3.

There exist multiple ways to back-propagate information from the venue level to the publication level depending on the available metadata:

1. based on the published venues (*Published-by*)
2. based on the referenced/cited venues (*References*)
3. based on the cited+citing venues (*References + Citations*)

***Published-by.*** (SciNoBo-pub): Given a publication $p$ and the set of distinct venues in which it has been published in $\{v_1, v_2, \cdots, v_n\}$ we draw weighted edges in layer $\ell_2$ of weight $w_{p,v} = \frac{1}{n}$. Thus, the weight is equally distributed among all published venues. Consequently, each venue contributes a score $\frac{w_{venue,FoS}}{n}$ to the publication. The scores per FoS are aggregated and ranked according to their total weights. The publication is finally classified to the top $T$ FoS, where $T$ might be fixed or be equal to the number of weights that exceed a user-defined threshold.

[14] https://www.oecd.org/sti/inno/frascati-manual.htm

[15] https://science-metrix.com/

***References.*** (SciNoBo-ref): Given a publication $p$ and a set of venues it references $\{v_1, v_2, \cdots, v_n\}$ we draw edges in layer $\ell_2$ of weight $w_{p,v}$ = (number of referenced publications published in $v$)/$n$. Each venue contributes a score $(w_{p,v})\cdot(w_{venue,FoS})$ where $w_{venue,FoS}$ is the normalized weight of the venue-FoS edge in $\ell 4$. Similar to the (*published-by approach*), the weights are aggregated and the publications are assigned to the top $T$ FoS.

***References + Citations.*** (SciNoBo-citref): This approach is identical to the reference approach except that in addition to referenced venues, the cited-by (citation) venues are also included in the venue set if available. A methodology originally proposed in the context of one particular field might eventually prove groundbreaking in a completely different field. By incorporating citation venues, SciNoBo captures cross-domain FoS that would otherwise be missed.

# 5 EXPERIMENTS

## 5.1 Dataset

In our experiments, we utilize the sciencemetrix Journal classification. sciencemetrix provides a list of Journals alongside with FoS labels (4.2). Furthermore, these FoS categories have been mapped to Level-3 FoS categories in our taxonomy (4.2).

To create, a comprehensive, large-scale, and clean dataset, we retrieve publications from Microsoft Academic Graph (mag) that are published in the Journals classified from sciencemetrix. mag provides a wide range of publications. Figure 3 presents the number of Journals that sciencemetrix has classified to Level-3 FoS in our taxonomy. One can easily observe, that by retrieving a certain amount of publications per journal, an unbalanced dataset will be created. We retrieve 500 publications per Journal and per Level 3 FoS. The unbalanced dataset created, describes real-world data, hence our evaluation splits follow this unbalanced distribution.

Moreover, we compare SciNoBo to a deep learning method, which requires a balanced train dataset. To that end, we further sample mag to obtain 10K train samples per Level 3 FoS. The final dataset statistics are presented in Table 3. For every publication, we also retrieve abstracts and additional metadata.

One limitation of the above-mentioned approach, is that the sciencemetrix classification is at the journal-level and not at the publication-level. We try to mitigate this effect by excluding multidisciplinary (e.g plos one) journals and assume that the journal-level classification also represents the publication-level classification.

**Table 3: Statistics of the dataset used for training and evaluating our methods.**

| Statistics | Train Set | Development Set | Test Set | Total |
|---|---|---|---|---|
| Number of Instances | 1687826 | 120282 | 120307 | 1928415 |

## 5.2 Baseline Method

We compare SciNoBo, to dann [21], a state-of-the-art model which utilizes textual information (abstracts) with a deep attentive neural network. dann [21], represents each abstract, by sorting the words based on tf-idf scores, keeping the top $d$ words, reverting to the original ordering and encoding them with pretrained Word Embeddings. Then, a deep neural encoder, with a bidirectional rnn and self-attention, encodes the abstract and a final softmax layer outputs the probabilities for each FoS category.

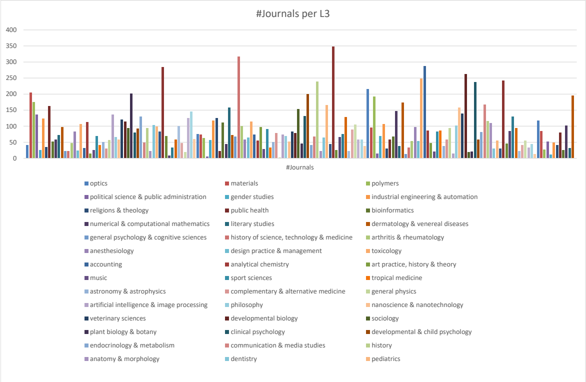

**Figure 3: Distribution of Journals, classified from sciencemetrix into Level-3 FoS categories in our taxonomy. Names of all Level 3 FoS labels were omitted for simplicity.**

Despite being one of the first works that utilized Deep Neural Networks, along with textual information to classify publications into FoS categories, the tf-idf filtering of the words of the abstracts breaks the sequence order and limits the effects of the rnn encoders. Furthermore, dann cannot perform hierarchical classification, whereas our approach is inherently hierarchical, since the SciNoBo network can provide publication-FoS codes at all 3 Levels of the taxonomy by employing the label propagation mechanism described in 4.3 and 4.4.

## 5.3 Implementation Details

Following [21], we employ their best model, which utilizes pretrained fasttext embeddings on a corpus created from WoS and tf-idf filtering of the abstracts by keeping the $d = 80$ most relevant words. To that end, we created a large corpus by retrieving publications from mag, assigned to Level 3 FoS labels in our taxonomy. The created corpus contains approximately 20 million publications with abstracts. We trained fasttext embeddings and calculated tf-idf scores. Regarding, the neural encoder of dann, we employed our pretrained fasttext embeddings with dimensionality of 100 and utilized 128 neurons at the recurrent encoders. Dropout was set at 0.2 and adam was used as the optimizer with learning rate of $10^{-3}$. We utilize Early Stopping with patience of 20, perform multiple experiments to account for standard deviation and present averaged results.

Regarding SciNoBo, no hyperparameter selection is required. After the Label Propagation (4.3) procedure is finished, the only information needed to infer publications to FoS categories, are published and citing/cited venues as stated in 4.4. We perform 2 rounds of Label Propagation.

**Table 4: macro-f1, weighted-macro-f1 and micro-f1 scores. Best scores are shown in bold. dann's experiments were repeated 4 times and averaged results are presented with standard deviation. The postfix top-1 and top-2 refer to the two settings described in Section 5.4.**

| models | macro-f1 | weighted-macro-f1 | micro-f1 |
|---|---|---|---|
| dann-top-1 | 0,4465 ± 0,003 | 0,5873 ± 0,002 | 0,4563 ± 0,002 |
| dann-top-2 | 0,6007 ± 0,008 | 0,7307 ± 0,002 | 0,6154 ± 0,003 |
| SciNoBo-citref-top-1 | 0,4627 ± 0,0 | 0,4900 ± 0,0 | 0,4900 ± 0,0 |
| SciNoBo-citref-top-2 | 0.6000 ± 0,0 | 0,6208 ± 0,0 | 0,6177 ± 0,0 |
| SciNoBo-ref-top-1 | 0,4781 ± 0,0 | 0,5000 ± 0,0 | 0,5022 ± 0,0 |
| SciNoBo-ref-top-2 | 0,6190 ± 0,0 | 0,6277 ± 0,0 | 0,6205 ± 0,0 |
| SciNoBo-pub-top-1 | **0,7303 ± 0,0** | **0,7309 ± 0,0** | **0,7503 ± 0,0** |
| SciNoBo-pub-top-2 | **0,8200 ± 0,0** | **0,8223 ± 0,0** | **0,8270 ± 0,0** |

## 5.4 FoS Classification & Evaluation

Given that dann (5.2) cannot perform hierarchical classification, evaluation was carried out at level-L3 of our classification scheme (i.e., 174 FoS Labels). In Section 4.4, the different approaches of classifying publications with SciNoBo have been explored. To analyze the impact of each classification approach, we present results for each variant. Evidently, this analysis can also be viewed as an ablation study.

The evaluation dataset and the baseline (dann) support only multiclass classification, whereas SciNoBo can be utilized in multi-label and multiclass tasks. Publication-level classification cannot always be addressed as a multiclass task, since a growing number of multidisciplinary publications is published, and journals slowly shift towards that direction. We perform multiclass evaluation to be aligned with the created dataset and baseline, but also present results with two settings to cater for multidisciplinarity. top-1 where we output only the most-probable FoS Label and top-2 where we output the two most-probable FoS Labels.

We compute macro-f1 and micro-f1 to compare performance between SciNoBo and dann. Recall that our test set is unbalanced (5.1) and to account for it we also compute weighted-macro-f1.

## 5.5 FoS Classification Results

Field of Science classification results are reported in Table. 4. Regarding macro-f1, weighted-macro-f1 and micro-f1 our variant of SciNoBo that utilizes only the published venues (SciNoBo-pub) outperforms all the other methods, in both evaluation settings (top-1, top-2). This presumably can be attributed to the nature of our evaluation dataset. Since the labeling of the publications in our dataset originates from labeling at the journal level, we can view SciNoBo-pub as a method to perform journal classification, in effect same as sciencemetrix. One would expect SciNoBo-pub to perform much better than already did. This is not the case, because we re-evaluate the initial venue label assignment during Label Propagation (4.3). This label re-assignment originates from the neighborhood structure of the venue in question, indicating that the original label assignment should be reconsidered or that by aggregating more than one FoS labels the venue leaned towards multidisciplinarity.

SciNoBo-ref clearly outperforms in macro-f1 and micro-f1 both SciNoBo-citref and dann in top-1 setting and still outperforms them in top-2 setting, however with the results being more competitive. We conclude that is important to take into account the references of a publication when it is published, which are always available (even on the first day of publication), unlike textual information. Furthermore, dann's performance might be hindered by only utilizing abstracts, since many FoS labels have overlapping terminology (3). On the other hand, this effect is mitigated in SciNoBo because authors usually cite similar work in their publications, making the FoS label more easily interpretable.

However, all of the methods perform much better in the top-2 setting, revealing that the correct FoS label (according to the dataset) is frequently in the second most probable position. This implies the multidisciplinary nature of publications and suggests the need for creating a multi-label publication-level dataset to account for multidisciplinarity, which we leave for future work.

For weighted-macro-f1, dann outperforms SciNoBo-citref and SciNoBo-ref in both settings, achieving a high score in the top-2 setting. Recall that our evaluation sets are unbalanced. weighted-macro-f1 assigns a weight to the FoS labels according to the number of samples that each FoS label has in the evaluation set. dann performs much better in this setting showing that it classifies correctly more high weighted FoS labels. Whereas our methods perform mostly similar in both metrics suggesting that presumably SciNoBo overall generalizes better but performs poorly in some high-weighted FoS labels.

One key observation, is that SciNoBo-ref performs slightly better than SciNoBo-citref in all three metrics and in both settings, suggesting that as time evolves and publications receive more and more citations their primary FoS label shifts thematically.

Finally, dann is a deep neural attentive network and apart from computational training time (23 hours per experiment) with each experiment, results slightly deviate. Furthermore, differences in hardware, limit the reproducibility of the results. SciNoBo has no computational/hardware overhead, apart from Label Propagation which is in the order of minutes and does not require a GPU, deviation of the results is non-existent which makes it more robust. The output will deviate only if the multilayer graph is changed or a publication receives more citations.

## 5.6 Qualitative analysis

To better understand the differences of each approach and to further establish the aforementioned arguments in Section 5.5, we present qualitative results.

**Inferring with SciNoBo-ref and SciNoBo-citref**: Table 5 presents three publications along with additional metadata (doi, title, abstract) and the inferred FoS label when inferring with SciNoBo-ref and SciNoBo-citref. We observe that SciNoBo-ref inferred all three publications successfully.

This behavior can be attributed to the fact that when a publication is published, it cites very similar content-wise publications. By examining the titles and abstracts, we can verify that the labeling of SciNoBo-ref is correct. However, it is evident that all three publications can be applied to other FoS as well. For example, the third publication examines an application of artificial intelligence

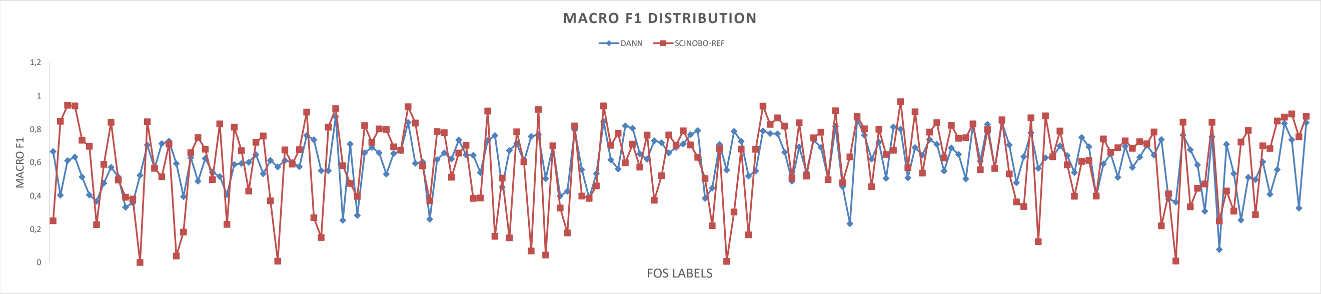

**Figure 4: MACRO-F1 distribution of all the FoS labels sorted by ascending order according to the number of instances. The names of the labels were omitted for simplicity.**

**Table 5: Publications presented with metadata (DOI, TITLE, ABSTRACT, PUBLISHED VENUE) along with the true FoS label of our dataset and the inferred FoS label of SciNoBo-REF and SciNoBo-CITREF. Only snippets of the abstracts are presented.**

| DOI | TITLE | PUBLISHED VENUE | SciNoBo-REF | SciNoBo-CITREF | LABEL |
|---|---|---|---|---|---|
| 10.3788/COL201715.062201 | Versatile nanosphere lithography technique combining multiple exposure nanosphere lens lithography and nanosphere template lithography | Chinese Optics Letters | Optics | NanoScience & Technology | Optics |
| 10.1089/ast.2019.2203 | The Role of Meteorite Impacts in the Origin of Life | Astrobiology | Astronomy & Astrophysics | Developmental Biology | Astronomy & Astrophysics |
| 10.1080/08839514.2018.1506971 | Strategic Particle Swarm Inertia Selection for Electricity Markets Participation Portfolio Optimization | Applied Artificial Intelligence | Artificial Intelligence & Image Processing | Energy | Artificial Intelligence & Image Processing |

in energy. As time evolves, the publication received citations from other works that are involved in energy and are published in energy venues. This behavior shows the shift of FoS labels in publications with time and shows the importance of creating a publication-level multi-labeled dataset. The above-mentioned argument explains the lower MACRO-F1 and MICRO-F1 scores in SciNoBo-CITREF, which according to the labeling of the dataset inferred the publications incorrectly.

**Case study for SciNoBo and DANN**: In Figure 4, we present the MACRO-F1 distribution of SciNoBo-REF and DANN for all the FoS labels, ordered by ascending order according to the number of instances. Recall that DANN only outperforms SciNoBo and its variants in the WEIGHTED-MACRO-F1 setting. Note that the MACRO-F1's scores remain the same in both settings but in WEIGHTED-MACRO-F1 they are weighted according to the number of instances (support) each FoS label has in the evaluation set. We observe that the overall performance of SciNoBo-REF is better than DANN. However, SciNoBo-REF performs poorly in some of the FoS labels, whereas DANN performs fairly well in them indicating the reason for the high WEIGHTED-MACRO-F1 score of DANN.

## 6 CONCLUSIONS AND FUTURE WORK

We propose SciNoBo, a new method to perform Field of Science (FoS) classification along with a new taxonomy based on the classification scheme of the OECD disciplines/fields of research and development (FORD) and SCIENCEMETRIX journal classification. SciNoBo along with the FoS taxonomy are inherently hierarchical and can perform multi-label and multiclass evaluations across all disciplines as opposed to previous work. Our proposed method leverages the strengths of utilizing minimal metadata that are always available, even at the first day a publication is published. By incorporating citing/publishing relationships into a Multilayer Graph containing publications-venues-FoS Labels, we are able to provide "context-aware" classifications without relying on the publication content as in many previous works. Furthermore, since our method can utilize citations that publications received, we can perform case studies showcasing the multidisciplinary nature of publications and how they can be assigned to more than one FoS labels in the course of time. We evaluated our method in a dataset created from SCIENCEMETRIX classification and MAG publications. Even though our dataset and baseline support only multiclass evaluation, experimental results and qualitative analysis demonstrated that our method is effective and outperforms a deep-learning method which rely solely on abstract information.

In future work, we plan to create a hierarchical publication-level multi-labeled dataset to better understand the benefits of SciNoBo and to encourage further research. In addition, we plan to extend our FoS taxonomy to broader levels, to provide a better sense of granularity, which will also help us to identify emerging FoS labels and classify publications to them. Moreover, we will explore alternative data sources, such as the OpenAlex[16] public repository, as well as, alternative ways of assessing relations between publications beyond direct citation scores.

## ACKNOWLEDGMENTS

The authors thank the three anonymous reviewers for their constructive feedback. The authors would also like to thank Dimitris Pappas for his invaluable support on data collection, conceptualization of the layer-layer inference, and proposal of the multiplication of normalized weights during the inference step. This work was supported by research grants from the European Union's Horizon 2020 Research and Innovation Programme under grant agreement

[16] https://docs.openalex.org/

101004870 (https://cordis.europa.eu/project/id/101004870) and the EOSC Future Project under grant agreement 101017536 (https://cordis.europa.eu/project/id/101017536).

## REFERENCES

[1] Mehwish Alam, Russa Biswas, Yiyi Chen, Danilo Dessì, Genet Asefa Gesese, Fabian Hoppe, and Harald Sack. 2021. HierClasSArt: Knowledge-Aware Hierarchical Classification of Scholarly Articles. In *Companion Proceedings of the Web Conference 2021*. Association for Computing Machinery, New York, NY, USA, 436–440. https://doi.org/10.1145/3442442.3451365

[2] Éric Archambault, Olivier H Beauchesne, and Julie Caruso. 2011. Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th international conference of the international society for scientometrics and informetrics*. Durban South Africa, 66–77.

[3] Jeroen Baas, Michiel Schotten, Andrew Plume, Grégoire Côté, and Reza Karimi. 2020. Scopus as a curated, high-quality bibliometric data source for academic research in quantitative science studies. *Quantitative Science Studies* 1, 1 (2020), 377–386. Publisher: MIT Press One Rogers Street, Cambridge.

[4] Caroline Birkle, David A Pendlebury, Joshua Schnell, and Jonathan Adams. 2020. Web of Science as a data source for research on scientific and scholarly activity. *Quantitative Science Studies* 1, 1 (2020), 363–376. Publisher: MIT Press One Rogers Street, Cambridge.

[5] Annette Boaz, Siobhan Fitzpatrick, and Ben Shaw. 2009. Assessing the impact of research on policy: A literature review. *Science and Public Policy* 36, 4 (May 2009), 255–270. https://doi.org/10.3152/030234209X436545 _eprint: https://academic.oup.com/spp/article-pdf/36/4/255/4693984/36-4-255.pdf.

[6] Lutz Bornmann, Robin Haunschild, and Rüdiger Mutz. 2021. Growth rates of modern science: a latent piecewise growth curve approach to model publication numbers from established and new literature databases. *Humanities and Social Sciences Communications* 8, 1 (2021), 1–15. Publisher: Palgrave.

[7] Cornelia Caragea, Florin Bulgarov, and Rada Mihalcea. 2015. Co-training for topic classification of scholarly data. In *Proceedings of the 2015 conference on empirical methods in natural language processing*. Association for Computational Linguistics, Lisbon, Portugal, 2357–2366.

[8] Worasit Choochaiwattana. 2010. Usage of tagging for research paper recommendation. In *2010 3rd International Conference on Advanced Computer Theory and Engineering (ICACTE)*, Vol. 2. IEEE, Chengdu, China, V2–439. https://doi.org/10.1109/ICACTE.2010.5579321

[9] Lisa Colledge. 2014. Snowball metrics recipe book. *Amsterdam: Snowball Metrics Program Partners* 110 (2014), 82.

[10] Lisa Colledge and R Verlinde. 2014. Scival metrics guidebook. *Netherlands: Elsevier* (2014), 68.

[11] Noshir Contractor, Peter Monge, and Paul M Leonardi. 2011. Network Theory| multidimensional networks and the dynamics of sociomateriality: bringing technology inside the network. *International Journal of Communication* 5 (2011), 39.

[12] Joshua Eykens, Raf Guns, and Tim C. E. Engels. 2021. Fine-grained classification of social science journal articles using textual data: A comparison of supervised machine learning approaches. *Quantitative Science Studies* 2, 1 (April 2021), 89–110. https://doi.org/10.1162/qss_a_00106 _eprint: https://direct.mit.edu/qss/article-pdf/2/1/89/1906557/qss_a_00106.pdf.

[13] Eugene Garfield, Morton V Malin, and Henry Small. 1975. A system for automatic classification of scientific literature. *Journal of the Indian Institute of Science* 57, 2 (1975), 14.

[14] Belver C Griffith, Henry G Small, Judith A Stonehill, and Sandra Dey. 1974. The structure of scientific literatures II: Toward a macro-and microstructure for science. *Science studies* 4, 4 (1974), 339–365. Publisher: Sage Publications Sage CA: Thousand Oaks, CA.

[15] Ioanna Grypari, Dimitris Pappas, Natalia Manola, and Haris Papageorgiou. 2020. Research & Innovation Activities' Impact Assessment: The Data4Impact System. In *Proceedings of the 1st Workshop on Language Technologies for Government and Public Administration (LT4Gov)*. European Language Resources Association, Marseille, France, 22–27. https://aclanthology.org/2020.lt4gov-1.4

[16] Zaynab Hammoud and Frank Kramer. 2020. Multilayer networks: aspects, implementations, and application in biomedicine. *Big Data Analytics* 5, 1 (2020), 1–18. Publisher: Springer.

[17] Qi He, Bi Chen, Jian Pei, Baojun Qiu, Prasenjit Mitra, and Lee Giles. 2009. Detecting Topic Evolution in Scientific Literature: How Can Citations Help?. In *Proceedings of the 18th ACM Conference on Information and Knowledge Management (CIKM '09)*. Association for Computing Machinery, New York, NY, USA, 957–966. https://doi.org/10.1145/1645953.1646076 event-place: Hong Kong, China.

[18] Ginny Hendricks, Dominika Tkaczyk, Jennifer Lin, and Patricia Feeney. 2020. Crossref: The sustainable source of community-owned scholarly metadata. *Quantitative Science Studies* 1, 1 (2020), 414–427. Publisher: MIT Press One Rogers Street, Cambridge.

[19] Christian Herzog, Daniel Hook, and Stacy Konkiel. 2020. Dimensions: Bringing down barriers between scientometricians and data. *Quantitative Science Studies* 1, 1 (2020), 387–395. Publisher: MIT Press One Rogers Street, Cambridge.

[20] Fabian Hoppe, Danilo Dessì, and Harald Sack. 2021. Deep Learning Meets Knowledge Graphs for Scholarly Data Classification. In *Companion Proceedings of the Web Conference 2021*. Association for Computing Machinery, New York, NY, USA, 417–421. https://doi.org/10.1145/3442442.3451361

[21] Bharath Kandimalla, Shaurya Rohatgi, Jian Wu, and C Lee Giles. 2021. Large scale subject category classification of scholarly papers with deep attentive neural networks. *Frontiers in research metrics and analytics* 5 (2021), 31. Publisher: Frontiers.

[22] Brigitte Khoury, Cary Kogan, and Sariah Daouk. 2017. International Classification of Diseases 11th Edition (ICD-11). In *Encyclopedia of Personality and Individual Differences*, Virgil Zeigler-Hill and Todd K. Shackelford (Eds.). Springer International Publishing, Cham, 1–6. https://doi.org/10.1007/978-3-319-28099-8_904-1

[23] Mikko Kivelä, Alex Arenas, Marc Barthelemy, James P Gleeson, Yamir Moreno, and Mason A Porter. 2014. Multilayer networks. *Journal of complex networks* 2, 3 (2014), 203–271. Publisher: Oxford University Press.

[24] Karol Kurach, Krzysztof Pawlowski, Lukasz Romaszko, Marcin Tatjewski, Andrzej Janusz, and Hung Son Nguyen. 2013. Multi-label Classification of Biomedical Articles. In *Intelligent Tools for Building a Scientific Information Platform: Advanced Architectures and Solutions*, Robert Bembenik, Lukasz Skonieczny, Henryk Rybinski, Marzena Kryszkiewicz, and Marek Niezgodka (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 199–214. https://doi.org/10.1007/978-3-642-35647-6_15

[25] Walter Leal Filho, Ulisses Azeiteiro, Fátima Alves, Paul Pace, Mark Mifsud, Luciana Brandli, Sandra S Caeiro, and Antje Disterheft. 2018. Reinvigorating the sustainable development research agenda: the role of the sustainable development goals (SDG). *International Journal of Sustainable Development & World Ecology* 25, 2 (2018), 131–142. Publisher: Taylor & Francis.

[26] Loet Leydesdorff and Ismael Rafols. 2009. A global map of science based on the ISI subject categories. *Journal of the American Society for Information Science and Technology* 60, 2 (2009), 348–362. Publisher: Wiley Online Library.

[27] Micha\l Lukasik, Tomasz Kuśmierczyk, \Lukasz Bolikowski, and Hung Son Nguyen. 2013. Hierarchical, multi-label classification of scholarly publications: modifications of ML-KNN algorithm. In *Intelligent tools for building a scientific information platform*. Springer, 343–363.

[28] Paolo Manghi, Alessia Bardi, Claudio Atzori, Miriam Baglioni, Natalia Manola, Jochen Schirrwagen, Pedro Principe, Michele Artini, Amelie Becker, Michele De Bonis, and others. 2019. The OpenAIRE research graph data model. *Zenodo* (2019), 23.

[29] OECD. 2015. *Frascati Manual 2015.* Organisation for Economic Cooperation and Development. https://www.oecd-ilibrary.org/content/publication/9789264239012-en

[30] Francesco Osborne, Angelo Salatino, Aliaksandr Birukou, and Enrico Motta. 2016. Automatic classification of springer nature proceedings with smart topic miner. In *International Semantic Web Conference*. Springer, 383–399.

[31] Antonio Perianes-Rodriguez and Javier Ruiz-Castillo. 2017. A comparison of the Web of Science and publication-level classification systems of science. *Journal of Informetrics* 11, 1 (2017), 32–45. Publisher: Elsevier.

[32] Silvio Peroni and David Shotton. 2020. OpenCitations, an infrastructure organization for open scholarship. *Quantitative Science Studies* 1, 1 (2020), 428–444. Publisher: MIT Press One Rogers Street, Cambridge.

[33] M Rivest, E Vignola-Gagne, and E Archambault. 2021. Article-level classification of scientific publications: A comparison of deep learning, direct citation and bibliographic coupling. *PloS one* 16, 5 (2021).

[34] Javier Ruiz-Castillo and Ludo Waltman. 2015. Field-normalized citation impact indicators using algorithmically constructed classification systems of science. *Journal of Informetrics* 9, 1 (2015), 102–117. Publisher: Elsevier.

[35] Angelo Salatino, Francesco Osborne, and Enrico Motta. 2021. CSO Classifier 3.0: a scalable unsupervised method for classifying documents in terms of research topics. *International Journal on Digital Libraries* (2021), 1–20. Publisher: Springer.

[36] Angelo Salatino, Thiviyan Thanapalasingam, Andrea Mannocci, Francesco Osborne, and Enrico Motta. 2018. Classifying Research Papers with the Computer Science Ontology. In *SEMWEB*.

[37] Angelo A Salatino, Thiviyan Thanapalasingam, Andrea Mannocci, Francesco Osborne, and Enrico Motta. 2018. The computer science ontology: a large-scale taxonomy of research areas. In *International Semantic Web Conference*. Springer, 187–205.

[38] Zhihong Shen, Hao Ma, and Kuansan Wang. 2018. A web-scale system for scientific knowledge exploration. *arXiv preprint arXiv:1805.12216* (2018).

[39] Fei Shu, Charles-Antoine Julien, Lin Zhang, Junping Qiu, Jing Zhang, and Vincent Larivière. 2019. Comparing journal and paper level classifications of science. *Journal of Informetrics* 13, 1 (2019), 202–225. Publisher: Elsevier.

[40] Henry Small, Kevin W Boyack, and Richard Klavans. 2014. Identifying emerging topics in science and technology. *Research policy* 43, 8 (2014), 1450–1467. Publisher: Elsevier.

[41] Vilius Stanciauskas, Ioanna Grypari, Gustaf Nelhans, G Papageorgiou, and I Demiros. 2020. Policy report on new indicators and approaches for assessing the

societal impact of re-search and innovation activities: Big Data approaches for improved monitoring of re-search and innovation performance and assessment of the societal impact in the Health, Demographic Change and Wellbeing Societal Challenge.

[42] Jie Tang, Jing Zhang, Limin Yao, Juanzi Li, Li Zhang, and Zhong Su. 2008. ArnetMiner: Extraction and Mining of Academic Social Networks. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD '08)*. Association for Computing Machinery, New York, NY, USA, 990–998. https://doi.org/10.1145/1401890.1402008 event-place: Las Vegas, Nevada, USA.

[43] S Upham and Henry Small. 2010. Emerging research fronts in science and technology: patterns of new knowledge development. *Scientometrics* 83, 1 (2010), 15–38. Publisher: Akadémiai Kiadó, co-published with Springer Science+ Business Media BV.

[44] Thanasis Vergoulis, Ilias Kanellos, Claudio Atzori, Andrea Mannocci, Serafeim Chatzopoulos, Sandro La Bruzzo, Natalia Manola, and Paolo Manghi. 2021. Bip! db: A dataset of impact measures for scientific publications. In *Companion Proceedings of the Web Conference 2021*. Association for Computing Machinery, 456–460.

[45] Thanasis Vergoulis, Ilias Kanellos, Serafeim Chatzopoulos, Christos Tryfonopoulos, Theodore Dalamagas, and Yannis Vassiliou. 2018. Pub Finder: Assisting the discovery of qualitative research. Association for Computing Machinery, Larnaka, Cyprus.

[46] Reinhilde Veugelers, Michele Cincera, Rainer Frietsch, Christian Rammer, Torben Schubert, Anita Pelle, Andrea Renda, Carlos Montalvo, and Jos Leijten. 2015. The impact of horizon 2020 on innovation in Europe. *Intereconomics* 50, 1 (2015), 4–30. Publisher: Springer.

[47] Ludo Waltman and Nees Jan Van Eck. 2012. A new methodology for constructing a publication-level classification system of science. *Journal of the American Society for Information Science and Technology* 63, 12 (2012), 2378–2392. Publisher: Wiley Online Library.

[48] Kuansan Wang, Zhihong Shen, Chiyuan Huang, Chieh-Han Wu, Yuxiao Dong, and Anshul Kanakia. 2020. Microsoft academic graph: When experts are not enough. *Quantitative Science Studies* 1, 1 (2020), 396–413. Publisher: MIT Press One Rogers Street, Cambridge.