Received: 12 November 2019 | Revised: 27 September 2020 | Accepted: 17 October 2020

DOI: 10.1002/asi.24433

RESEARCH ARTICLE

JASIST WILEY

Topic diversity: A discipline scheme-free diversity measurement for journals

Yi Bu | **Mengyang Li** | **Weiye Gu** | **Win-bin Huang**

Department of Information Management, Peking University, Beijing, China

Correspondence
Win-bin Huang, Department of Information Management, Peking University, Beijing 100871, China.
Email: huangwb@pku.edu.cn

Abstract

Scientometrics has many citation-based measurements for characterizing diversity, but most of these measurements depend on human-designed categories and the granularity of discipline classifications sometimes does not allow in-depth analysis. As such, the current paper proposes a new measurement for quantifying journals' diversity by utilizing the abstracts of scientific publications in journals, namely topic diversity (*TD*). Specifically, we apply a topic detection method to extract fine-grained topics, rather than disciplines, in journals and adapt certain diversity indicators to calculate *TD*. Since *TD* only needs as inputs abstracts of publications rather than citing relationships between publications, this measurement has the potential to be widely used in scientometrics.

1 | INTRODUCTION

Studies on interdisciplinary scientific research have been quite popular in scientometrics and social studies of science. Wagner et al. (2011) defined *interdisciplinarity* as integration of separate disciplinary data, methods, tools, concepts, and theories to create a holistic view or common understanding of a complex issue, question, or problem. To quantify interdisciplinarity, bibliometricians have built up many useful indicators or measurements, such as diversity (e.g., Leydesdorff, Wagner, & Bornmann, 2019), entropy (e.g., Stirling, 2007), and betweenness centrality (e.g., Leydesdorff, 2007). Among these, diversity is a concept from ecology by Stirling (2007) and Rafols and Meyer (2010) that takes three components into account, namely variety (number of categories), balance (relative number of elements in each category), and disparity (difference or similarity between categories).

One branch of studies on diversity aims to quantify the degree of interdisciplinarity of journals. Rousseau and Glanzel (2016), for instance, employed a new diversity measurement from Leinster and Cobbold (2012). They combined human-made discipline classification schema, citation analysis, and their diversity measurement to quantify journals' diversity. Leydesdorff, Wagner, and Bornmann (2018) ranked journals' interdisciplinarity by exploring different functions of betweenness centrality and diversity in various types of journal-level scholarly networks (e.g., citation, co-citation, and bibliographic coupling).

Although there are many existing studies focusing on diversity measurements in scientometrics (e.g., Leydesdorff, 2007; Leydesdorff, Kushnir, & Rafols, 2014; Leydesdorff & Rafols, 2011; Rafols & Meyer, 2010; Skupin, Biberstine, & Borner, 2013), "disciplines" and "subjects" in these indicators are mostly represented by manually assigned categories—for instance, Zhang et al. (2016) employed the Leuven-Budapest (ECOOM) subject-classification scheme. Yet, there are at least three limitations of adopting such human-assigned schemes for diversity measurements. First, these schemes are static so they could not reflect the dynamics of discipline or subject evolution and structures. For instance, Frank, Wang, Cebrian, and Rahwan (2019) observed that the field of artificial intelligence is being combined with

Yi Bu and Mengyang Li contributed equally to this study.

mathematics and computer science at an accelerating rate, indicating potential changes of discipline schemes. Second, as argued by Waltman (2016), discipline or subject classification is subjective and often fails to get a consensus from different perspectives. Third, the granularity of disciplines does not allow more in-depth analyses in, for instance, sub-fields or research topics. Take Zhang et al. (2016) as an example: Although they have examined the diversity at the journal level, they cannot easily dig into their knowledge structure (i.e., the structure and linkage at the level of research topics or sub-fields).

To this end, in this paper, we propose a new diversity measurement for journals that does not depend on any existing subject classification scheme, namely topic diversity (*TD*). This new measurement requires as inputs the abstracts of publications in a certain journal; its output is a real number that quantifies how diverse the research topics of this journal are. There are mainly four steps in the calculation of our proposed measurement, namely word extraction, network construction, topic detection, and diversity calculation. In the first and the second steps, we implement some basic natural language processing and select candidate words that are semantically "meaningful" to the following steps by considering the topological structure of the co-word network. In the third step, we detect candidate topics (communities in the co-word network) and filter them to obtain reasonable topics. In the last step, we calculate *TD* of each journal by considering variety, balance, and disparity.

There are three highlights of our proposed measurement:

1. This measurement does not depend on any predefined subject classification system, and it, therefore, reduces potential biases and potentially enables us to investigate the dynamics of subject or discipline developments;
2. This measurement indicates a more fine-grained diversity by considering topic-level, instead of purely subject-level, information; and.
3. This measurement does not require as inputs the details of citing or cited publications of the articles published in the targeted journal, which extends its application for future research.

2 | RELATED WORK

2.1 | Diversity measurements

The concept and measurement of diversity in applied statistics were borrowed from ecology. At its early stage, diversity only took variety and balance into account, such as Simpson diversity (mentioned in Jost, 2006). A decade ago, Stirling (2007) introduced the concept of diversity to the field of quantitative science studies and added disparity as another important element to characterize diversity. To avoid bias, Stirling assigned disparity and balance with different weights which could be adjusted to fit for various tasks. Specifically, he defined diversity as $D_{\alpha,\beta} = \sum_{\substack{i,j=1 \\ i \neq j}}^{N} d_{ij}^{\alpha} \left(p_i p_j\right)^{\beta}$ where N is the number of categories, p_i (p_j) is the ratio of items in category i (j), d_{ij} is the distance or dissimilarity of categories i and j, and α, β are two given positive parameters to adjust the importance (weight values) of d_{ij} and $p_i p_j$.

Jost (2009) later provided six requirements for a "true diversity measurement": symmetry, zero output independence, transfer principle, homogeneity, replication principle, and normalization. Unfortunately, the measurement proposed by Stirling (2007) does not satisfy all these requirements. Meanwhile, although Hill's diversity (1973) could fulfill these requirements, it does not take disparity into consideration. To propose a measurement that not only integrates variety, balance, and disparity but also satisfies the principles proposed by Jost (2009), Leinster and Cobbold (2012) put forward a complex formulation of diversity, $D_q^S = \left(\sum_{i=1}^{N} p_i \left(\sum_{j=1}^{N} s_{ij} p_j\right)^{q-1}\right)^{\frac{1}{1-q}}$, where s_{ij} is the similarity of categories i, j, and $s_{ii} = 1$ (q is a positive parameter). The range of D_q^S is between one (for a system that has only one category) and N (for a system that has N categories in which all categories have an equal number of items and each two of them are quite different in that their similarity is zero). Leydesdorff and Ivanova (in press) discussed recent advances in diversity as well as other interdisciplinarity measurements.

Diversity was quantitatively and systematically introduced into Information Science by Rafols and Meyer (2010). They built a conceptual framework to understand interdisciplinarity and knowledge integration in which there are two indicators, namely discipline diversity and coherence. Disciplinary diversity is used to understand the heterogeneity of a set of bibliometric entities (e.g., institute, journal, author, or even the discipline itself), while coherence measures the extent of similarity in a set of bibliometric entities. Based on Rafols and Meyer (2010), Liu, Rafols, and Rousseau (2012) introduced a more generalized theoretical framework. Particularly, they distinguished three types of entities: the *source set* (articles in certain discipline of journal) of enquiry, an *intermediary set* derived from the source, and a *target set* (categories). These entities consist of two mapping

relations (i.e., from source set to intermediary set, and from intermediary set to target set) which reveal knowledge integration and knowledge diffusion, as the maps could lead the knowledge in the entities of the source set into various entities in the intermediary set and finally merge in different categories. Most subsequent diversity measurements under the context of Information Science accord with this framework (i.e., source, intermediary, and target sets), including our proposed measurement.

The indicator proposed by this paper, *TD*, is based on the framework of Liu et al. (2012). Our *source set* contains the abstracts of scientific publications in a certain journal. The *intermediary set* contains keywords extracted from the abstracts in the source set. The *target set* is research topics, which are a group of keywords and are clustered from the co-word network constructed by the keywords and their co-occurrence relations from all journals. Co-occurrence relations of words have been widely adopted in scientometrics as a useful way of studying the content of publications. Leydesdorff and Hellsten (2006), for instance, employed word co-occurrence relations to indicate the changes of the meanings of some selected words between documents in different contexts. Rokaya, Atlam, Fuketa, Dorji, and Aoe (2008) combined word co-occurrence relations with tf-idf to improve the performance of information retrieval. Hellsten and Leydesdorff (2020) investigated the social media data by researching the co-occurrence of actors and topics to find the topic that gained more attention online.

2.2 | Top-down versus bottom-up approaches

As pointed out by Liu et al. (2012), there are several issues worth noting when calculating diversity. The first issue is whether the indicator adopts a top-down (categories assigned by human) or bottom-up (categories automatically clustered) approach. Top-down approaches are more popular (e.g., Carley & Porter, 2012; Leydesdorff et al., 2018; Leydesdorff & Rafols, 2011; Porter & Rafols, 2009; Zhang et al., 2016) as most bibliographic databases assigned at least one category to each paper (e.g., Web of Science [WoS] and Microsoft Academic Graph [MAG]). However, top-down methods are limited in representing real-world discipline or research topic. Bottom-up approaches, on the other hand, tend to mine the inner structure of scholarly networks and extract clusters as disciplines. This branch of methods (e.g., Leydesdorff et al., 2014; Shen, Chen, Yang, & Wu, 2019; Skupin et al., 2013) is more suitable for detecting emergent fields that change frequently. In our proposed indicator, we adopt the bottom-up approach to extract research topics from co-word networks to calculate the diversity of journals. Meanwhile, in our indicator, different from top-down approaches, the intermediary set and target set are automatically extracted and fit for any given source set (publications). In the current paper, we apply the D_2^S "true" diversity formula (Zhang et al. (2016)) to calculate the value of *TD* as well as another formula, *DIV* (Leydesdorff et al., 2019), which was proposed recently for interdisciplinarity of publications. Although Leydesdorff et al. (2019) adopted the total number of categories (N) as the number of journals in the same discipline, we argue that it does not fit for the meaning of diversity, as diversity could reveal interdisciplinarity, so its variety ought to be based on all scientific disciplines. One solution to deal with the scope of *DIV* variety is put forward by Rousseau, Zhang, and Hu (2019) that removes N from the *DIV* variety formula. In this paper, we remain N so the value is linear to the formula of Rousseau et al. (2019) and does not affect the *ranking order* of each journal. Meanwhile, our *DIV* formula can still be limited between zero and one while the diversity value of Rousseau's revised formula has no upper limit. The details of D_2^S and *DIV* are discussed in the Methodology section.

2.3 | Quantifying "distance" of categories

Another important issue is quantifying the "distance" of categories, which is often difficult to implement manually. Regardless of top-down or bottom-up methods, the distance is basically measured by structural similarity of scholarly networks. In detail, each network could be represented as an adjacent matrix, where each row or column is a certain bibliometric entity (e.g., a discipline, a journal, a paper, etc.) and cells show the relation or link or interaction between the corresponding two entities (a corresponding row and a column in the matrix). Therefore, the similarity between two disciplines could be quantified by some commonly used measurements, for example, cosine similarity. Details of these measurements and their potential applications (scopes) have been discussed in White (2003).

As for human-given categories like WoS classification, Leydesdorff, Carley, and Rafols (2013) built a global map of journals and calculated the similarity of each discipline. Meanwhile, if the discipline is automatically clustered (Jensen & Lutkouskaya, 2014), the similarity is still calculated from the publication network. Once the similarity s is acquired, the distance d could be defined as $1 - s$ (Leydesdorff et al., 2013) or $1/s$ (Jensen & Lutkouskaya, 2014). Particularly, the first one limits d between zero and one and thus fits for more general circumstances that require normalization, and the distance is linearly correlated with similarity.

3 | METHODOLOGY

The abstract of an article usually represents the emphasis of the work and contributions. Words adopted in the abstract reveal the core content of a scientific publication. As a result, to measure the *TD* of a journal, we select the abstracts of all publications in the journal as the input of the proposed algorithm. The procedure of the algorithm, shown in Figure 1, comprises four steps, namely word extraction, network construction, topic detection, and diversity calculation. In the first step, some meaningful words from all the selected scientific articles in the journal are obtained by the word extractor from the abstracts. After all publications in the journal have been processed, the overall extracted words are collected and connected by our network constructor as a large word co-occurrence network of all disciplines. The topic detector is then exploited to group the relevant words with strong connections in the network, and each obtained word group is regarded as a relevant scientific topic. Finally, the *TD* of the journal is calculated by the diversity calculator according to the words from the publications in the journal, and each word could be either meaningless or refer to one topic. The detail of the algorithm in each step is described in the following sections.

3.1 | Step 1: Word extraction

The objective of this step is to obtain a set of meaningful words that can represent scientific knowledge in one article. The word extractor, E, acquiring these words from an article, is defined in Equation (1).

$$E(\boldsymbol{T_{mi}}) = \boldsymbol{D}_{mi}, m \in [1,N], i \in [1,M_i] \quad (1)$$

where a total of N journals are included and M_i articles exist in a certain journal, m, and $\boldsymbol{D}_{mi}$ (which contains elements in the *intermediary set*) is the set of the meaningful words extracted from the text of abstract $\boldsymbol{T}_{mi}$ (which is one element of the *source set*) of the ith article of the journal m. Assume that the number of words in the article is $|\boldsymbol{D}_{mi}|$, and $\boldsymbol{D}_{mi} = \left\{V_{mi1}, V_{mi2}, ..., V_{mij}, ..., V_{mi|\boldsymbol{D}_{ij}|}\right\}$.

In the word extractor, as shown in Figure 2, the abstract of the publication should be processed through a three-phase procedure in the extractor, namely text preprocessing, word selection, and phrase extraction.

3.1.1 | Text preprocessing

In the text preprocessing phase, basic text processing techniques are exploited to obtain all candidate meaningful words of the article. Tokenizing the abstract is first executed to obtain single words rather than phrases for further usage. Part-of-speech tagging is then applied to find a particular part of speech corresponding to each word in these texts. After that, words which are not nouns or adjectives are discarded. The remainder is then lemmatized and the letters are transformed to lowercase. Finally, the output words,

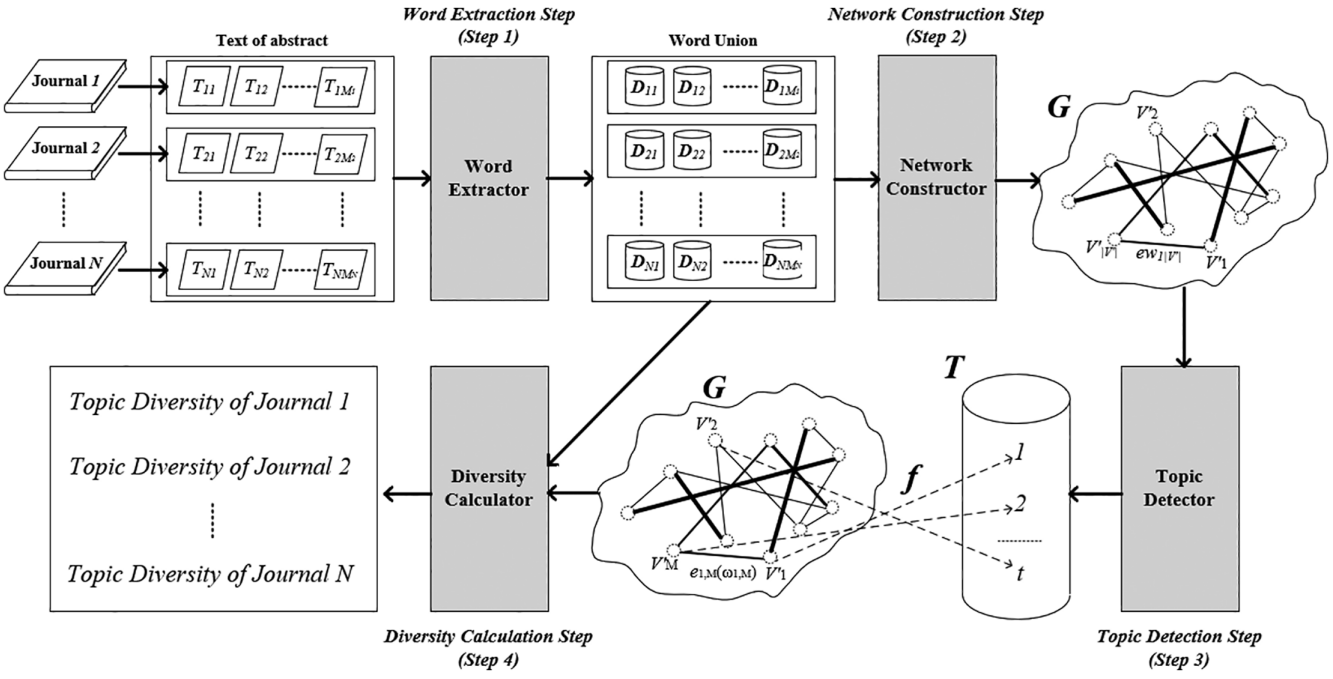

FIGURE 1 The procedure of the proposed algorithm

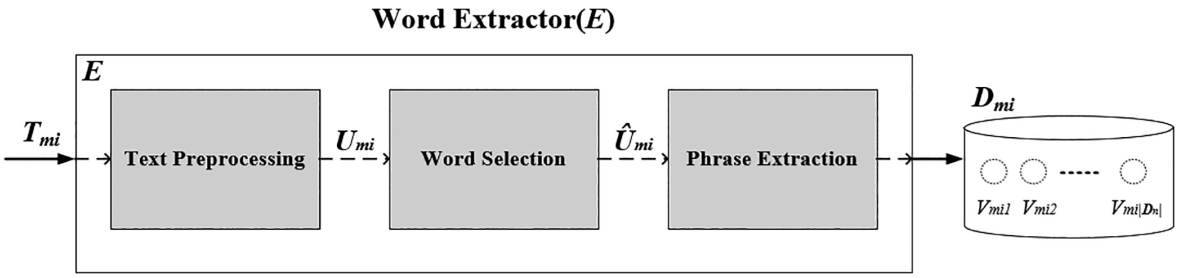

FIGURE 2 The block diagram of the word extractor

$\boldsymbol{U}_{mi} = \{U_{mi1}, U_{mi2}, ..., U_{mij}, ...\}$, in this phase are obtained after we remove general stop words and common words.

3.1.2 | Word selection

In the word selection phase, a graph-based ranking algorithm inspired by PageRank, called TextRank (Mihalcea & Tarau, 2004), is mainly adopted to select the meaningful words in the text from the candidates. The purpose of TextRank is to rank words by constructing an undirected graph of which the nodes are the candidate words in the text preprocessing phase obtained from all publications (i.e., the whole word union $\boldsymbol{W} = (\boldsymbol{U}_{11} \cup ... \cup \boldsymbol{U}_{1M_1}) \cup (\boldsymbol{U}_{21} \cup ... \cup \boldsymbol{U}_{2M_2}) \cup ... \cup (\boldsymbol{U}_{N1} \cup ... \cup \boldsymbol{U}_{NM_N})$) and the edges are the relations between two such words. In the current context, we consider the co-occurrence relationship in TextRank, a commonly used strategy in bibliometrics (e.g., Yan & Ding, 2012); specifically, two nodes (words) are linked in the graph if they co-occur within a certain window size in the original text. The weight of an edge equals the number of co-occurrences between the two words (nodes). Normally, the window size ranges from 2 to 10 words and it is suggested to set it as two to reach the best performance, according to Mihalcea and Tarau (2004). We construct a co-occurrence network and calculate the TextRank score for each node (word). The TextRank score, $S(\boldsymbol{W_i})$, of each node in $\boldsymbol{W}$ is calculated iteratively by using Equation (2):

$$S(\boldsymbol{W_i}) = (1-d) + d\sum_{\boldsymbol{W_j} \text{ connected to } \boldsymbol{W_j}} \frac{w_{ij}}{\sum_{\boldsymbol{W_k} \text{ connected to } \boldsymbol{W_j}} w_{jk}} S(\boldsymbol{W_j}) \quad (2)$$

where d is a damping factor between zero and one, usually set to 0.85 (Brin & Page, 1998). w_{ij} is the number of co-occurrences between two words, $\boldsymbol{W_i}$ and $\boldsymbol{W_j}$, in the given time window. After all the words in $\boldsymbol{W}$ are processed, the words that rank top third or above in terms of their scores are selected into $\hat{\boldsymbol{W}}$ as the output in the word selection phase. Meanwhile, for each $\boldsymbol{U}_{mi}$, if the word U_{mij} is not included in $\hat{\boldsymbol{W}}$, it will be removed from $\boldsymbol{U}_{mi}$. In the end, only those "significant" words with high PageRank values in $\hat{\boldsymbol{W}}$ will remain in the word set of each article, which is stored as $\hat{\boldsymbol{U}}_{mi}$.

3.1.3 | Phrase extraction

In the phrase extraction phase, if some selected words in $\hat{\boldsymbol{U}}_{mi}$ are adjacent in the original text (i.e., abstract), we will count them (two or more words) as one phrase, and replace the single words with a new phrase (e.g., "information" and "retrieval" often occur adjacently, and we will consider "information retrieval" instead of separately counting in the next steps). Then, each $\hat{\boldsymbol{U}}_{mi}$ is transformed into $\boldsymbol{D}_{mi}$, which includes the set of meaningful words and phrases as the final output of the word extractor.

3.2 | Step 2: Network construction

In this step, we exploit a network constructor to assemble an undirected weighted co-word network, $\boldsymbol{G}$, which covers important and meaningful words that well indicate the major semantic information of all publications with minimal redundancy. Two phases, namely word union and edge connection, are included in the network constructor, as shown in Figure 3.

3.2.1 | Word union

After all the articles have been processed by the word extractor in Step 1, their word sets, $\boldsymbol{D}_{mi}$, are gathered and joined into a big word set as $\boldsymbol{V'} = \{V'_1, V'_2, ..., V'_i, ..., V'_{|\boldsymbol{V'}|}\} \subseteq \bigcup_{mi} \boldsymbol{D}_{mi}$ in the word union phase, where each component in $\boldsymbol{V'}$ (a.k.a., $V'_1, V'_2, ..., V'_i, ..., V'_{|\boldsymbol{V'}|}$) represents a set of words with the

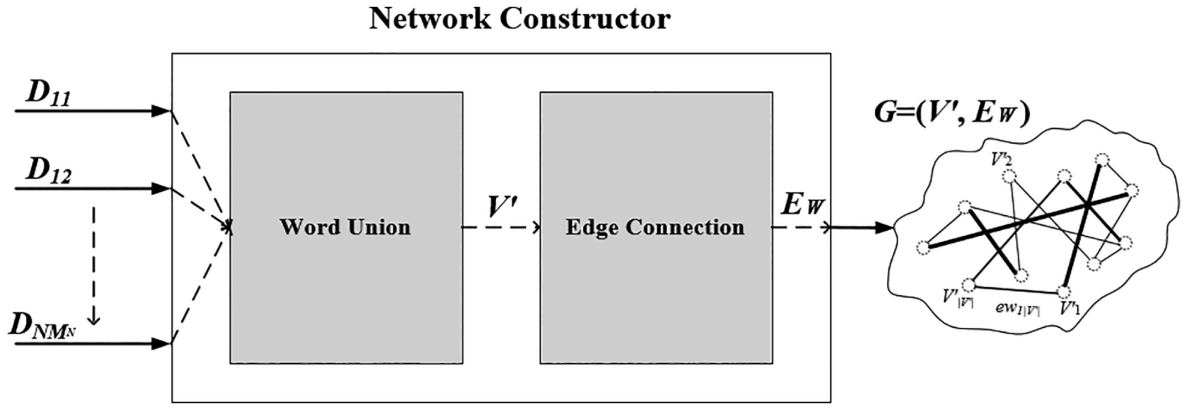

FIGURE 3 The block diagram of network construction

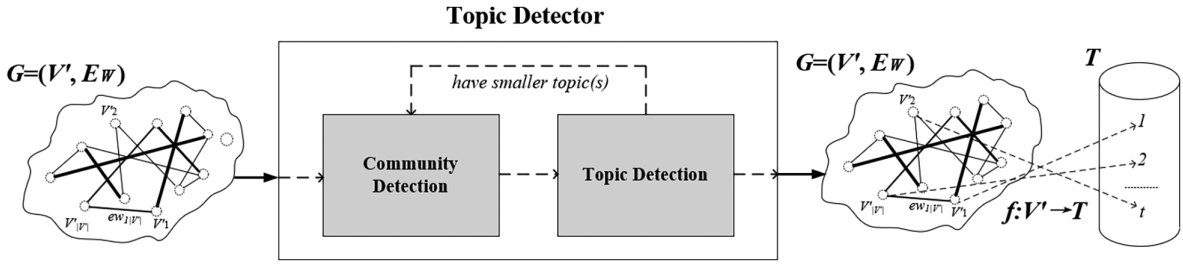

FIGURE 4 The block diagram of the topic detector

same meaning, that is, they are synonyms or cognate words. The current paper adopts all scientific disciplines in the empirical study. Thus, we can hardly make a global thesaurus to group synonyms. To this end, we have to only group the words (or phrases) that have the same stem (as for phrases, they should have the same stem for each word contained in a given phrase). $|\boldsymbol{V}'| \leq \sum_{ij} |\boldsymbol{D}_{mi}|$ is the number of words in it. Here, V'_i with its synset should be different from $V'_{j \neq i}$.

3.2.2 | Edge connection

In the edge connection phase, two nodes, V'_i and V'_j, in $\boldsymbol{V}'$ are connected if they co-occurred in an article of the dataset, and the edge, ew_{ij}, is weighed as their co-occurrence frequency. After all the edges, $\boldsymbol{Ew} = \{ew_{11}, ew_{12}, ..., ew_{ij}, ...\}$, are connected and calculated, the network, $\boldsymbol{G} = (\boldsymbol{V}', \boldsymbol{Ew})$, is constructed.

3.3 | Step 3: Topic detection

In this step, we exploit a topic detector to identify the research topics of a journal, as shown in Figure 4.

Particularly, we developed an algorithm based on the Louvain method (Blondel, Guillaume, Lambiotte, & Lefebvre, 2008) as the strategy to detect potential fine-grained (candidate) topics. The input of the topic detection algorithm is the network (**G**) generated from Step 2, while the output is a mapping function $f: \mathbf{V}' \rightarrow \mathbf{T}$, where $\mathbf{V}'$ is the node set of **G** and $\mathbf{T} = \{1, 2, ..., t\}$ is the topic index set. In this way, f attaches each node to one certain candidate topic. Details of the Louvain method can be found in Appendix A. However, the Louvain method typically outputs tens of communities; thus, the number of nodes in one community could be averagely quite huge and there are often thousands of words in one community. Therefore, in our method, we recursively adopt the Louvain method to obtain "deeper" (in terms of content) and smaller sub-topics. Specifically, we annotate the first group of topics (i.e., the first round of output of the Louvain method) as T0, and extract sub-graphs of the original **G** while each sub-graph contains all the words in one T0 topic. We then employ the Louvain method again with these sub-graphs to get more sub-topics, which are called T1. In this way, we could continue to acquire T2, T3, T4, etc. If, by implementing the Louvain method in a certain Ti ($i > 0$) topic, we cannot get any "meaningful" topic (the criterion for distinguishing meaningless communities or topics is discussed in the next paragraph), then we stop the

recursion. With this algorithm, **G** could be divided thoroughly into fine-grained communities to get specific topics.

Although one graph or sub-graph might be divided into several small communities, some of them (usually comprised of only a few words [i.e., nodes]) may not have "actual" meaning. To this end, we define a criterion to filter communities detected by the recursive Louvain method based on the topology of communities. Specifically, for a certain node $\boldsymbol{\xi}$, let $\mathbf{k}_{\xi}$ be the degree within the community (only neighbors in the same community are counted towards its degree). $\bar{\mathbf{k}}$ and $\mathbf{S_k}$ are the mean and the standard deviation of $\mathbf{k}_{\xi}$, respectively. We expect to select nodes with a great value of z-score ($=\frac{\mathbf{k}_{\xi}-\bar{\mathbf{k}}}{\mathbf{S_k}}$) as hubs in their communities. If a community has no hubs (i.e., nodes in this community all have similar degrees), we argue that this community is unrepresentative for the topic, because "real" topics making more sense tend to contain some core terms. In practice, we follow Guimera and Amaral (2005), who set 2.5 as the threshold of nodes' *z*-scores to detect hubs in a community.

In this step, we obtain a mapping function, $f': \mathbf{V}' \rightarrow \mathbf{T}$, where $\mathbf{T} = \{1, 2, ..., t\}(t \leq c)$ (the target set) is the updated community index set (essentially the topic index set, as now each filtered community represents a reasonable topic). Our method assigns each word to only one topic. Although in real scientific disciplines, one word or phrase may have multiple meanings and thus be assigned to different topics, this will certainly affect the content of topics but not the results of diversity measurements. This is because, if a few words are falsely assigned to different wrong topics, we can imagine that the diversity results may be barely affected as their ratio is extremely small compared with other correctly assigned keywords. On the other hand, if there are plenty of words assigned to one wrong topic, as the topics are grouped by keywords' co-occurrence relations, this topic must have a strong linkage with the "correct" topic, and the two topics are quite likely to be similar and basically do not affect final diversity (because their disparity is low). In either case, the final diversity value will not be distorted too much.

3.4 | Step 4: Diversity calculation

In this step, we exploit a diversity calculator to quantify the TD of a journal, as shown in Figure 5.

In this paper, we put forward two formulas for the diversity calculator: D_q^S and DIV. D_q^S is a diversity measurement detailed by Leinster and Cobbold (2012) and we follow Zhang et al., 2016 to set q = 2. The corresponding TD measurement, namely $\mathbf{D_2^S}$**-TD**, of journal m is defined as:

$$\mathbf{D_2^S - TD_m} = \left(\sum_{i=1}^{t} p_{mi} \sum_{j=1}^{t} s_{ij} p_{mj}\right)^{-1} = \frac{1}{\sum_{i,j=1}^{t} s_{ij} p_{mi} p_{mj}} \quad (3)$$

where $p_{mi} = \frac{\sum_{f(v)=i, v \in \mathbf{D_m}} 1}{|\mathbf{D_m}|}$ and s_{ij} is calculated from the co-word network, **G**, and word-topic mapping function, $f': \mathbf{V}' \rightarrow \mathbf{T}$. We quantify the relation of two topics, i, j, as $R_{ij} = \sum_{f(v)=i, f(u)=j, v,u \in V} ew_{vu} (i \neq j)$ while $R_{ii} = 0$. The cosine similarity of topics i and j equals $s_{ij} = \frac{\sum_{k=1}^{t} R_{ik} R_{jk}}{\sqrt{\left(\sum_{k=1}^{t} R_{ik}^2\right)\left(\sum_{k=1}^{t} R_{jk}^2\right)}}$.

Leinster and Cobbold (2012) argued that this measurement takes into consideration all of the three dimensions of diversity. First of all, it involves the t, the number of topics, as a parameter, indicating variety. Second, p_{mi} de facto indicates the ratio of topic i in journal m; thus, the variance of p_{m1}, p_{m2}, ..., and p_{mt} preliminarily illustrates balance. Last, s_{ij} as the similarity measurement quantifies how close different topics are, demonstrating disparity (Zhang et al. (2016)). Mathematically, the range of D_2^S-TD is between one and t. Other

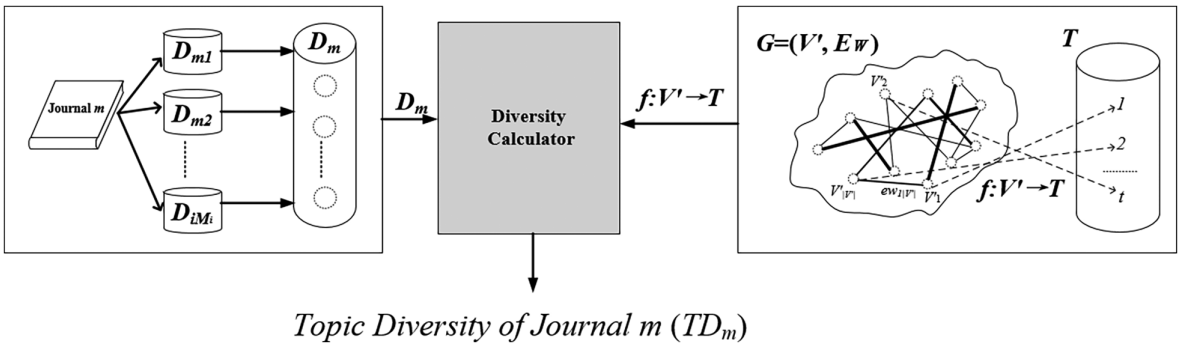

FIGURE 5 The block diagram of the diversity calculator

mathematical properties of this indicator have been discussed by Leinster and Cobbold (2012) and summarized by Zhang et al. (2016).

Meanwhile, another state-of-the-art diversity formula, DIV (Leydesdorff et al., 2019), is also suitable for the diversity calculator to adapt to. Typically, the DIV-TD is calculated as:

$$\mathrm{DIV-TD_m} = \frac{n_m}{N} \times (1 - \mathrm{Gini}) \times \sum_{i,j,i \neq j}^{n_m} \frac{d_{ij}}{n_m(n_m - 1)} \quad (4)$$

Equation (4) multiplies variety ($\frac{n_m}{N}$, n_m is the number of categories (topics) in journal m, and N is the total number of categories (topics)), balance (Gini index indicates unbalance), and disparity (d_{ij} refers to disparity of categories i and j, and $n_m(n_m - 1)$ is used for normalization so that disparity could be limited between zero and one). Leydesdorff et al. (2019) observed that DIV outperforms Rao-Stirling diversity and betweenness centrality, and that DIV could intuitively reveal variety, balance, and disparity.

4 | EMPIRICAL STUDIES

4.1 | Data and processing

In the empirical study, we employ the Microsoft Academic Graph (MAG) dataset (Sinha et al., 2015) that covers scientific publications, their citing relations, metadata of publications, their authorship information, publication venue, and other related information. We first select all publications in 2005 and 2006 recorded in MAG as the candidate publications to be used in further analyses in the current paper. For each publication, MAG labels one or more discipline(s) and the labeled discipline(s) are at different levels (i.e., L0–L5, where L0 contains the most macro-level disciplines). Following Alshebli, Rahwan, and Woon (2018), we select L0 to identify a publication's discipline; there are in total 19 disciplines in L0 (Table 1).

The distinct number of journals for these 2005–2006 publications recorded in MAG equals 29,365. Additionally, we also assign a journal's "discipline" by examining its publications' disciplines: Specifically, the discipline with the maximum number of publications in the journal will be assigned to the journal as its "discipline." In this way, each journal is assigned to one specific L0 discipline.

Due to the extreme computational complexity, we have to select only some journals, instead of all, from all 29,365 journals. For each L0 discipline, the number of journals we select from this discipline is proportional to its share of journals in all L0 disciplines. This is to make sure the sampled journal-discipline distribution is the same as the real one. When selecting journals from a specific discipline, we first rank all journals in this discipline by their 2007 impact factors; the impact factors are calculated based on the MAG dataset. Next, the top 20% of journals in each discipline with the highest impact factors are sampled. After collecting the top 20% of journals, we only keep journals that meet the following four conditions simultaneously: (a) The journal contains at least 10 publications in 2005–2006; (b) the journal has at least received one citation in 2007[1]; (c) publications (2005–2006) in the journal are written in English; and (d) at least one publication (2005–2006) in the journal has its abstract recorded.[2] These journals are called *candidate journals* in our empirical study. Finally, 3,444 journals are selected. Table 1 shows the number and percentage of journals in each L0 discipline. After selecting

TABLE 1 Number of selected journals and number of publications in each L0 discipline

Discipline	Number of journals	Number of publications
Medicine	696	223,821
Biology	374	136,966
Engineering	342	42,603
Sociology	305	18,949
Psychology	290	43,651
Economics	282	22,630
Mathematics	201	40,672
Computer Science	198	20,244
Chemistry	196	120,845
Political Science	139	5,849
History	118	5,966
Business	113	50,324
Materials Science	112	8,090
Geology	94	29,533
Physics	86	100,327
Philosophy	55	3,575
Art	51	2,473
Environmental Science	37	2,970
Geography	17	1,525
Total	3,444	881,013

Note: The percentage of journals in each discipline is proportional to its share of journal counts among all samples. The disciplines here are ranked by their number of journals.

the 3,444 candidate journals, we collect the abstracts from all the publications in the selected journals, as well as their citing and cited publications. We will use these data to calculate *TD* and other indicators for comparison.

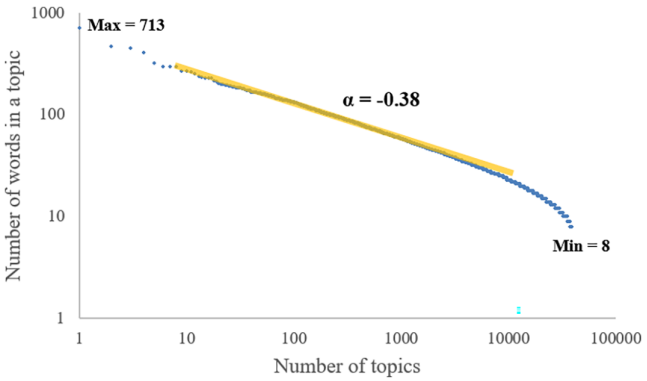

FIGURE 6 The distribution of the size of topics ($N = 38{,}855$) [Color figure can be viewed at wileyonlinelibrary.com]

After Word Extraction and Network Construction steps, we extract 2,306,866 unique keywords (single words or phrases, words or phrases with the same stems are grouped into one keywords). By implementing our topic detection algorithm with abstracts from the publications of the 3,444 journals, we identify 38,855 topics. These topics cover 794,857 keywords, while the rest of the keywords are filtered as meaningless words which do not belong to any topic (detailed in "Step 3: Topic Detection" section). The distribution of the size (number of words on one topic) of 38,855 topics is shown in Figure 6. We can see that these data points form an approximately straight line in a double-logarithmic coordinate, which indicates that the size of topics follows a power-law distribution. We also observe that most topics contain fewer than 100 words.

TABLE 2 Top 20 topics with their ID, size (i.e., the number of words under this topic), and hubs

No.	ID[a]	Size	Hubs
1	0-12	713	Douglas/prestige/simon/close relationship
2	1-18	468	Universite
3	2-0-4-1-0-0-0-0-0-0	452	P/effect/activity/level/response/mechanism
4	1-2-2-2-0-2-0-0-0-0	408	System/paper/application/technique/generation
5	0-1-31	320	Internationalization/macro/web design/singh/future direction/proactive
6	1-3-0-0-0-1-0-0-0-0	297	Theory/term/point/limit
7	0-28	296	Malaria transmission
8	0-5-19	295	Specific condition/round/panelist
9	4-1-4-0-3-0	272	©/john
10	4-0-0-2-2-6-0	267	Period/J appl polym
11	3-7-1-0-1-0-1-0-1-0-1-0-0	262	Patient/disease/outcome/diagnosis
12	2-15	255	Cas-fed rat
13	0-0-4-3	237	Tractarianism/christina rossetti/oxford movement/keble/confession/faith/metaphor/rossetti
14	0-6-12	236	General education/family medicine clerkship/preceptor/medical student education/clinical education/predoctoral education
15	2-0-4-1-1-1-0-0	230	Role/important role/pathogenesis
16	3-2-3-19	228	Electronic database/available evidence/high quality rcts/search strategy
17	1-0-0-5-0-1-1-0-0-0-0	228	Data/time/comparison/event
18	0-5-0-26	216	Electronic health record/ehrs/health record/reimbursement
19	1-6-1-0-0-1-0	214	Field/magnetic field/electric field
20	3-2-1-1-1-0-0-0-0-0-0-0	205	Use/rate/investigation

[a]The format is T0-T1-T2-...T*n*, while T*i* is the community detected by the *i*th round Louvain Method in the topic detection step.

TABLE 3 Six diversity indicators

Dimensions: Raw data type versus diversity formulas	Abstract text	Citing relation	Cited relation
$D_2^S = \sum_{ij} \frac{1}{s_{ij} p_i p_j}$	***D_2^S-TD***	D_2^S-citing	D_2^S-cited
$DIV = \frac{n}{N} \times (1 - Gini) \times \sum_{i,j,i \neq j}^{n} \frac{d_{ij}}{n(n-1)}$	***DIV-TD***	*DIV*-citing	*DIV*-cited

Note: All these indicators are summarized in the gray shading area. Each indicator can be categorized based upon two dimensions, namely their raw data types and diversity formulas.

Table 2 presents the size and hubs (words with *z*-score more than 2.5, as aforementioned) of the top 20 topics. While it is difficult to interpret or explain some words (e.g., "universite" in the second topic. It might be a false stem from the word "university"), we can still find many meaningful topics. For instance, the 3rd, 4th, 6th, and 17th topics are fundamental topics of many scientific works, reflected by some representative words such as "system" and "theory". The 5th and 16th topics concern computer science and the 7th, 8th, 11th, 12th, 14th, 15th, and 18th might belong to medical-related topics. Meanwhile, the 15th and 18th topics seem to be close to education and health informatics. The large proportion of topics that are related to medicine and biological topics is consistent with the huge amount of journals or publications in these domains (Bu, 2020). The details of all 38,855 topics are listed in Appendix B.

4.2 | Indicators for comparison

In addition to D_2^S-*TD* and *DIV-TD*, we also calculate some other indicators of characterizing diversity for a better comparison. All these diversity indicators could be classified based upon two dimensions, namely raw data type and diversity formulation. In terms of the *raw data type*, while *TD* adopts text data (abstract) for measuring diversity, existing methods tend to employ citing relations between publications in journals. For these methods involving citing relations, they consider both citing and cited sides' details (Leydesdorff et al., 2019). The second dimension is *diversity formulas*, for which we select two state-of-the-art formulas: D_2^S (Zhang et al. (2016)) and *DIV* (Leydesdorff et al., 2019), as mentioned in Step 4 in the Methodology.

Table 3 summarizes all of the six diversity indicators adopted in this paper. All these indicators are summarized in the gray shading area of Table 3. Each indicator can be categorized based upon two dimensions, namely their raw data types and diversity formulas. For instance, the diversity indicator D_2^S-citing can be viewed from two perspectives. In terms of the raw data type, it adopts

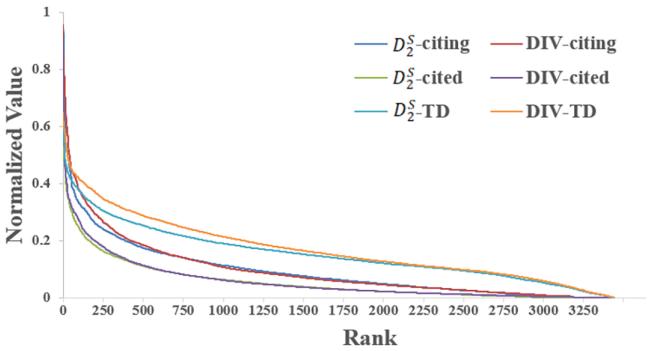

FIGURE 7 The distribution of the six diversity indicators. All values are normalized in [0,1] [Color figure can be viewed at wileyonlinelibrary.com]

citing relations of publications; as for the diversity formula, it uses D_2^S (Zhang et al. (2016)). In Table 3, our proposed *TD* indicators are in *bold*; the row and column indicate that *TD* employs abstract texts of publications and adopts D_2^S and *DIV*.

More specifically, these indicators are calculated as the following steps:

D_2^S-*TD*: For this indicator, we collect the abstracts of the publications in each journal, and follow the Methodology section in this paper to calculate the journal's value of diversity using the D_2^S formula. Due to the computational complexity, we only sample 2,658 topics which contain more than 40 words. They cover 21.11% of the total number of keywords (167,789/794,857). The other 36,197 smaller topics are filtered as meaningless communities.
DIV-TD: For this indicator, we collect the abstracts of the publications in each journal, and follow the Methodology section in this paper to calculate the journal's value of diversity using the *DIV* formula. Due to the computational complexity, we only sample 2,658 topics which contain more than 40 words.D_2^S-citing: For this indicator, we collect the citing publications of the publications in each journal. Only citing publications within the 3,444 journals are considered. We adopt $D_2^S = \sum_{ij} \frac{1}{s_{ij} p_i p_j}$ (Zhang et al. (2016)) to calculate the D_2^S-citing diversity of each

TABLE 4 Statistics of the six indicators

Indicator	Mean	Median	Max.	Min.	*N*
D_2^S-*TD*	84.58318	76.09582	545.4849	1	3,444
DIV-*TD*	0.040928	0.035445	0.241123	0	3,444
D_2^S-citing	8.197372	6.000000	79.76203	1	3,444
DIV-citing	0.003749	0.002291	0.039746	0	3,444
D_2^S-cited	8.580131	5.235879	138.5243	1	3,444
DIV-cited	0.003981	0.002047	0.069686	0	3,444

TABLE 5 Top 20 journals of D_2^S-*TD* and *DIV*-*TD*. Both diversity indicators are based on abstract text

Rank	Journal	D_2^S-*TD*	Journal	*DIV*-*TD*
1	Proceedings of the National Academy of Sciences of the United States of America	545.4849	Proceedings of the National Academy of Sciences of the United States of America	0.2411
2	Biochemical and Biophysical Research Communications	336.7656	Biochemical and Biophysical Research Communications	0.1567
3	Geophysical Research Letters	330.4889	Biochimica et Biophysica Acta	0.1496
4	Analytical Chemistry	321.9105	Geophysical Research Letters	0.1488
5	Biochimica et Biophysica Acta	318.1066	Journal of the American Chemical Society	0.1486
6	Journal of Physical Chemistry B	317.8577	Cancer Research	0.1478
7	Journal of Applied Physics	312.9445	Journal of Physical Chemistry B	0.1442
8	Journal of the American Chemical Society	300.2726	Analytical Chemistry	0.1426
9	Applied Physics Letters	287.7782	Chemosphere	0.1420
10	Journal of Nutrition	284.8396	Journal of Clinical Oncology	0.1403
11	Electrophoresis	283.2802	World Journal of Gastroenterology	0.1389
12	Kidney International	271.7072	Journal of Applied Physics	0.1388
13	Journal of Chemical Physics	267.5020	Brain Research	0.1386
14	Forensic Science International	264.1672	Journal of Geophysical Research	0.1383
15	Journal of Applied Physiology	262.9971	Physical Review E	0.1382
16	International Journal of Cancer	258.3699	Journal of Agricultural and Food Chemistry	0.1377
17	The New England Journal of Medicine	257.7832	Kidney International	0.1360
18	World Journal of Gastroenterology	256.7686	Journal of Chemical Physics	0.1338
19	Physica A Statistical Mechanics and Its Applications	255.4356	Applied Physics Letters	0.1317
20	Journal of Geophysical Research	255.0557	Transplantation Proceedings	0.1304

journal, while p_i refers to the proportion of journal i in all citing publications of the focal journal and s_{ij} refers to the similarity of journals i, j defined as the cosine similarity of the citation matrix.

DIV-citing: For this indicator, we collect the citing publications of the publications in each journal. Yet, different from D_2^S -citing, we employ $DIV = \frac{n}{N} \times (1 - Gini) \times \sum_{i,j,i \neq j}^{n} \frac{d_{ij}}{n(n-1)}$ to calculate the *DIV*-citing diversity of each journal, while n refers to the number of citing journals, N refers to total number of journals (=3,444), *Gini* is implemented using the proportion of all citing journals, and $d_{ij} = 1 - s_{ij}$.

D_2^S-*cited*: For this indicator, we collect the cited publications (until the end of 2016) of the publications in each journal. We utilize $D_2^S = \sum_{ij} \frac{1}{s_{ij} p_i p_j}$ to calculate the D_2^S-cited diversity of each journal, while p_i refers to the proportion of journal i in all cited publications of the focal journal, and s_{ij} refers to the cosine similarity of journals i, j.*DIV*-*cited*: For this indicator, we collect the cited publications (until the end of 2016) of the publications in

TABLE 6 Top 20 journals of D_2^S-citing and *DIV*-citing. Both diversity indicators are based on citing relations

Rank	Journal	D_2^S -citing	Journal	*DIV*-citing[a]
1	Pediatric Clinics of North America	79.7620	Proceedings of the National Academy of Sciences of the United States of America	0.0397 (0.0421)
2	Current Opinion in Clinical Nutrition and Metabolic Care	76.3094	Biochemical and Biophysical Research Communications	0.0380 (0.0329)
3	Pharmacology & Therapeutics	75.7637	Pharmacology & Therapeutics	0.0347 (0.0284)
4	Experimental Biology and Medicine	73.5539	Biochimica et Biophysica Acta	0.0330 (–[b])
5	Basic & Clinical Pharmacology & Toxicology	65.2174	The Scientific World Journal	0.0321 (–)
6	Expert Opinion on Drug Metabolism & Toxicology	64.7236	Expert Opinion on Pharmacotherapy	0.0321 (0.0257)
7	Current Opinion in Pharmacology	63.6445	Expert Opinion on Investigational Drugs	0.0311 (0.0204)
8	Annals of Medicine	63.2997	Cochrane Database of Systematic Reviews	0.0309 (0.0453)
9	BMC Public Health	60.0984	Current Opinion in Clinical Nutrition and Metabolic Care	0.0308 (0.0112)
10	Current Topics in Medicinal Chemistry	57.6261	World Journal of Gastroenterology	0.0302 (0.0281)
11	Lymphatic Research and Biology	50.6826	Journal of Pharmacology and Experimental Therapeutics	0.0297 (0.0192)
12	Histology and Histopathology	49.0645	Expert Opinion on Therapeutic Targets	0.0286 (0.0222)
13	Nature Reviews Drug Discovery	48.8395	Expert Opinion on Emerging Drugs	0.0279 (0.0100)
14	Clinical Therapeutics	48.2146	European Journal of Pharmacology	0.0278 (0.0307)
15	Expert Reviews in Molecular Medicine	47.4173	Brain Research	0.0276 (0.0216)
16	Current Opinion in Psychiatry	47.1895	Nature Reviews Drug Discovery	0.0274 (0.0142)
17	Journal of Pharmacology and Experimental Therapeutics	45.4986	BMC Public Health	0.0273 (0.0236)
18	Trends in Pharmacological Sciences	44.1276	Experimental Biology and Medicine	0.0265 (0.0340)
19	Clinical and Experimental Pharmacology and Physiology	43.0381	Expert Opinion on Drug Metabolism & Toxicology	0.0257 (0.0247)
20	Expert Opinion on Investigational Drugs	42.7882	Current Opinion in Pharmacology	0.0250 (0.0151)

[a]As citation-based results are calculated based on "local" citations among these 3,444 journals instead of the whole journal set, we provide the *DIV*-citing result of Leydesdorff et al. (2019) for comparison. They use the whole journal set of JCR (11,487 journals in 2016) to calculate citation-based DIV indicators. The result of Leydesdorff et al. (2019) is listed after our result and enclosed in brackets. Basically, our result and their result do not differ greatly.
[b]"–" indicates that this journal is not included in the 11,487 journals of JCR in its 2016 version.

each journal. Different from D_2^S -cited, we use $DIV = \frac{n}{N} \times (1 - Gini) \times \sum_{i,j,i \neq j}^{n} \frac{d_{ij}}{n(n-1)}$ to calculate the *DIV*-cited diversity of each journal.

4.3 | Results and discussions

Table 4 presents the descriptive statistics of the six diversity indicators. The rows are divided into three parts. The first part corresponds to our proposed indicators and shows two formulas of *TD*. The second part corresponds to citing-based indicators and presents two formulas of diversity. The third part corresponds to cited-based indicators and contains two formulas of diversity. From Table 4, we can see that the values of D_2^S -based indicators are all greater than one. Yet, the values of *DIV*-based indicators are much smaller. Furthermore, we observe that the value of *DIV-TD* is basically 10 times greater than those of *DIV*-citing and *DIV*-cited indicators, while the value of D_2^S-*TD* is 10 times greater than D_2^S -citing and D_2^S -cited indicators as well. This is attributable to the fact that the number of "categories" (journals) we select includes all scientific fields and that it is quite uncommon for one specific journal to cite or to be cited by a majority of existing scientific fields—this leads to a relatively low variety. This preliminarily indicates that text content (e.g., abstract in our experiment) may lead to a greater

TABLE 7 Top 20 journals of D_2^S-cited and *DIV*-cited. Both diversity indicators are based on cited relations

Rank	Journal	D_2^S -cited	Journal	*DIV*-cited[a]
1	Annual Review of Medicine	138.5243	The New England Journal of Medicine	0.0697 (0.1075)
2	Frontiers in Bioscience	102.5486	Proceedings of the National Academy of Sciences of the United States of America	0.0623 (0.1143)
3	Current Opinion in Pharmacology	91.7507	Biochemical and Biophysical Research Communications	0.0470 (0.0733)
4	Pharmacology & Therapeutics	73.7668	Cochrane Database of Systematic Reviews	0.0456 (0.0807)
5	Annals of Medicine	70.2973	Journal of Clinical Investigation	0.0441 (0.0674)
6	Current Opinion in Investigational Drugs	68.2721	Annals of Internal Medicine	0.0428 (0.0834)
7	Annual Review of Public Health	67.9010	Annual Review of Medicine	0.0410 (0.0394)
8	Annual Review of Psychology	67.3491	Frontiers in Bioscience	0.0375 (0.0418)
9	Expert Opinion on Therapeutic Targets	66.2988	International Journal of Cancer	0.0367 (0.0470)
10	Nature Reviews Drug Discovery	66.1301	Biochimica et Biophysica Acta	0.0331 (–)
11	International Journal of Oncology	65.2665	American Journal of Pathology	0.0324 (0.0580)
12	Annual Review of Pharmacology and Toxicology	64.8736	American Journal of Epidemiology	0.0323 (0.0643)
13	Physiological Reviews	64.1717	Nature Reviews Drug Discovery	0.0321 (0.0465)
14	Expert Opinion on Investigational Drugs	61.3126	Human Molecular Genetics	0.0321 (0.0423)
15	Current Opinion in Oncology	60.3215	Cancer Research	0.0319 (0.0488)
16	Expert Opinion on Pharmacotherapy	57.3888	Nature Reviews Genetics	0.0317 (0.0440)
17	Archives of General Psychiatry	56.0213	Journal of Cellular Physiology	0.0307 (0.0412)
18	The New England Journal of Medicine	55.8963	British Journal of Cancer	0.0303 (0.0497)
19	Cochrane Database of Systematic Reviews	55.7754	The American Journal of Medicine	0.0295 (0.0769)
20	Journal of Molecular Medicine	55.5691	Physiological Reviews	0.0291 (0.0589)

[a]We provide the *DIV*-cited result of Leydesdorff et al. (2019) for comparison. The result of Leydesdorff et al. (2019) is listed after our result and enclosed in brackets. Basically, their result is overall a little higher than ours. There are two possible reasons: they cover the whole corpus so that the publications could have more citation, or in 2016, journals are cited more than they are cited in 2005 because of citation inflation.

TABLE 8 Spearman rank-order correlation of the six diversity indicators ($N = 3{,}444$). All correlations are significant at the 0.01 level

	D_2^S-citing	*DIV*-citing	D_2^S-cited	*DIV*-cited	D_2^S-*TD*	*DIV-TD*	Impact factor
D_2^S-citing	–						
DIV-citing	0.892	–					
D_2^S-cited	0.724	0.716	–				
DIV-cited	0.681	0.817	0.920	–			
D_2^S-*TD*	0.491	0.685	0.562	0.720	–		
DIV-TD	0.458	0.685	0.548	0.732	0.933	–	
Impact factor	0.524	0.650	0.631	0.714	0.457	0.480	–

diversity value than citation-based information by considering more fine-grained topics which have more relations with each other through the co-word network. Since the original values of D_2^S and *DIV* differ greatly, we implement a min-max normalization strategy (i.e., $x_{normalized} = \frac{x_{original} - Min}{Max - Min}$) so that the diversity measurements are intuitively comparable without varying their rank. The distribution of the six diversity indicators is shown in Figure 7.

Tables 5–7 display the top 20 journals for each of the six diversity indicators. On the one hand, the top journals under D_2^S -*TD* and *DIV-TD* have more intersections (13 journals) than D_2^S-citing and *DIV*-citing (9 journals), as well as D_2^S-cited and *DIV*-cited (6 journals), indicators.

Although some journals rank in the top 20 under D_2^S-*TD* but not under *DIV-TD* (or rank in the top 20 under *DIV-TD* but not under D_2^S-*TD*), our result shows that they rank at least 69th (the D_2^S-*TD* rank of *Journal of Clinical Oncology*) in terms of *DIV-TD*/D_2^S-*TD*. Nonetheless, as for the two citation-based indicators, some of the top 20 journals fail to maintain a high diversity in terms of the other indicators. For instance, *Proceedings of the National Academy of Sciences of the United States of America* (PNAS) is listed as the top *DIV*-citing journal but its D_2^S-citing rank is only 359th. This is consistent with the result of Leydesdorff et al. (2019) although they only calculated citation-based indicators (i.e., Rao-Stirling diversity and *DIV*) but not text-based indicators. Therefore, we presume that text-based diversity indicators (*TD*) tend to have more consistent results when adopting a different diversity formula (D_2^S vs. *DIV*). We will discuss this further when performing the correlation analysis. On the other hand, although there are some overlaps among *TD* (Table 4), citing- (Table 5), and cited-based diversity (Table 6) for the top journal results (e.g., *Proceedings of the National Academy of Sciences of the United States of America*, *Biochemical and Biophysical Research Communications*, and *Nature Reviews Drug Discovery*), most of the top 20 *TD* journals do not have a very high citation-based diversity. We argue that these three types of diversity measurements may provide different aspects (text content, citing relations, and cited relations, respectively) to examine diversity. In terms of *TD*, we find many top-ranked journals in the fields of specific natural science disciplines, such as geophysics (e.g., *Geophysical Research Letters*, *Journal of Geophysical Research*), applied physics (e.g., *Journal of Applied Physics*, *Applied Physics Letters*), cancer (e.g., *Cancer Research*, *Journal of Clinical Oncology*) and chemistry (e.g., *Journal of the American Chemical Society*, *Analytical Chemistry*). *TD* reveals that these journals tend to be more topically diverse. As for citation-based indicators, there are many journals starting with "Current Opinion..." or "Expert Opinion..." (e.g., *Current Opinion in Clinical Nutrition and Metabolic Care*, *Current Opinion in Psychiatry*, *Expert Opinion on Emerging Drugs*, and *Expert Reviews in Molecular Medicine*). Publications in these journals tend to receive more citations or cite more references when offering opinions on a certain research topic (Waltman, 2016). Indeed, citation-based indicators focus more on the relation between different journals rather than their content. We argue that, when quantifying the diversity of journals, different aspects should be considered to obtain a comprehensive evaluation.

Table 8 shows the Spearman rank-order correlations among the six diversity indicators. We adopt the Spearman instead of Pearson's correlation because: (a) The indicators do not follow a typical normal distribution; and (b) we carry out a sampling of journals (3,000+ journals out of 20,000+) but the Spearman correlation will not be affected by sampling. From the table, we can find that D_2^S-*TD* has the highest correlation with *DIV-TD* among all the indicators (0.933). This accords with the results of the top 20 journals. The correlation coefficients between D_2^S-citing and *DIV*-citing (0.892) and between D_2^S- and *DIV*-cited (0.920) are much higher than for the others. This indicates that, with the same data type (e.g., citing, cited, or text data), D_2^S and *DIV* do not affect the final diversity value too much. On the other hand, the correlations between each two of *TD*, citing-, or cited-based indicators are much lower than the correlations between indicators having the same raw data type. This indicates that *TD*, citing-, or cited-based indicators might reveal different aspects of diversity. Nevertheless, the lowest Spearman correlation in Table 8 is 0.458 (between D_2^S-citing and *DIV-TD*), which means that citation- and text-based indicators still have some consistency. As for journal impact factor, among correlations between impact factor and the six diversity indicators, *DIV*-cited is the highest while *DIV-TD* is the lowest. This is reasonable as impact factor is based on citing relations rather than text content.

5 | CONCLUSIONS

This paper proposes an indicator to quantify the diversity of journals, called *TD*. *TD* does not depend on any predefined subject classification system. Thus, it reduces potential biases and potentially enables us to investigate the dynamics of subject or discipline developments. It indicates a more fine-grained diversity by considering topic-level instead of purely subject-level information. Meanwhile, *TD* does not require as inputs the details of citing or cited publications of the articles published in the targeted journal, which extends its application for future research. The empirical study shows that our proposed *TD* reveals additional information for diversity when compared with existing indicators such as *DIV*-related ones.

However, in the current definition, topics are derived from publication abstracts; some state-of-the-art natural language processing algorithms to extract keywords and detect topics might increase the performance of the indicator. It would also be quite interesting to employ a well-structured thesaurus or ontology to unify keywords. Furthermore, in this paper, we only select publications in English to acquire keywords. If the raw data is multilingual, it might be hard to extract keywords and topics in a common sense manner as there is basically no relation between terms or ideas across publications in different

languages—unless there are certain automatic machine translation techniques embedded in the algorithm.

In our experiment, due to the computational complexity, we only select citations among the selected 3,444 journals rather than all scientific journals, and we only sample 2,658 "larger" topics rather than all 38,855 topics. Although the result of the sample may approximate the whole corpus, it would still be better to use the whole corpus for a comprehensive empirical analysis. A future study could select all journals and all topics for a more reliable and credible result.

ACKNOWLEDGMENTS

This article was financially supported by Chinese National Funding of Social Sciences (No: 20BTQ054). The authors are grateful for fruitful discussions with Lin Zhang, Loet Leydesdorff, Caroline S. Wagner, and Lutz Bornmann. The authors would like to thank two anonymous reviewers for their insightful suggestions.

ORCID

Yi Bu https://orcid.org/0000-0003-2549-4580
Mengyang Li https://orcid.org/0000-0003-0489-2846

ENDNOTES

[1] The reason why at least one citation is required for this step is for convenience of calculating citing- or cited-based indicators as a comparison.

[2] The reason why at least one abstract is required for this step is for convenience of calculating *TD*. Some publications may not have an abstract or the abstract might not be included in the MAG corpus.

REFERENCES

AlShebli, B. K., Rahwan, T., & Woon, W. L. (2018). The preeminence of ethnic diversity in scientific collaboration. *Nature Communications*, *9*(1), 1–10.

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), 155–168.

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual web search engine. *Computer Networks and ISDN Systems*, *30*(1–7), 107–117.

Bu, Y. (2020). *Understanding the citation-based impact of scientific publications through ego-centered citation networks*. Indiana University Ph.D. thesis.

Carley, S., & Porter, A. L. (2012). A forward diversity index. *Scientometrics*, *90*(2), 407–427.

Frank, M. R., Wang, D., Cebrian, M., & Rahwan, I. (2019). The evolution of citation graphs in artificial intelligence research. *Nature Machine Intelligence*, *1*(2), 79–85.

Guimera, R., & Amaral, L. A. N. (2005). Cartography of complex networks: Modules and universal roles. *Journal of Statistical Mechanics: Theory and Experiment*, *13*(2), P02001.

Hellsten, I., & Leydesdorff, L. (2020). Automated analysis of actor–topic networks on twitter: New approaches to the analysis of socio-semantic networks. *Journal of the Association for Information Science and Technology*, *71*(1), 3–15.

Jensen, P., & Lutkouskaya, K. (2014). The many dimensions of laboratories' interdisciplinarity. *Scientometrics*, *98*(1), 619–631.

Jost, L. (2006). Entropy and diversity. *Oikos*, *113*(2), 363–375.

Jost, L. (2009). Mismeasuring biological diversity: Response to Hoffmann and Hoffmann (2008). *Ecological Economics*, *68*(4), 925–928.

Leinster, T., & Cobbold, C. A. (2012). Measuring diversity: The importance of species similarity. *Ecology*, *93*(3), 477–489.

Leydesdorff, L. (2007). Betweenness centrality as an indicator of the interdisciplinarity of scientific journals. *Journal of the American Society for Information Science and Technology*, *58*(9), 1303–1319.

Leydesdorff, L., Carley, S., & Rafols, I. (2013). Global maps of science based on the new web-of-science categories. *Scientometrics*, *94*(2), 589–593.

Leydesdorff, L., & Hellsten, I. (2006). Measuring the meaning of words in contexts: An automated analysis of controversies about 'monarch butterflies', 'frankenfoods', and 'stem cells'. *Scientometrics*, *67*(2), 231–258.

Leydesdorff, L., & Ivanova, I. (In Press). The measurement of 'Interdisciplinarity' and 'Synergy' in scientific and extra-scientific collaborations. *Journal of the Association Society for Information Science and Technology*, https://ssrn.com/abstract=3560339. https://doi.org/10.2139/ssrn.3560339

Leydesdorff, L., Kushnir, D., & Rafols, I. (2014). Interactive overlay maps for US patent (USPTO) data based on international patent classification (IPC). *Scientometrics*, *98*(3), 1583–1599.

Leydesdorff, L., & Rafols, I. (2011). Local emergence and global diffusion of research technologies: An exploration of patterns of network formation. *Journal of the American Society for Information Science and Technology*, *62*(5), 846–860.

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2018). Betweenness and diversity in journal citation networks as measures of interdisciplinarity—A tribute to Eugene Garfield. *Scientometrics*, *114*(2), 567–592.

Leydesdorff, L., Wagner, C. S., & Bornmann, L. (2019). Interdisciplinarity as diversity in citation patterns among journals: Rao-Stirling diversity, relative variety, and the Gini coefficient. *Journal of Informetrics*, *13*(1), 255–269.

Liu, Y., Rafols, I., & Rousseau, R. (2012). A framework for knowledge integration and diffusion. *Journal of Documentation*, *68*(1), 31–44.

Mihalcea, R., & Tarau, P. (2004). TextRank: Bringing order into text. In *Proceedings of the 2004 conference on empirical methods in natural language processing*, pp. 404–411.

Newman, M. E. J., & Girvan, M. (2004). Finding and evaluating community structure in networks. *Physical Review E*, *69*(2), 026113.

Porter, A. L., & Rafols, I. (2009). Is science becoming more interdisciplinary? Measuring and mapping six research fields over time. *Scientometrics*, *81*(3), 719–745.

Rafols, I., & Meyer, M. (2010). Diversity and network coherence as indicators of interdisciplinarity: Case studies in bionanoscience. *Scientometrics*, *82*(2), 263–287.

Rokaya, M., Atlam, E., Fuketa, M., Dorji, T. C., & Aoe, J. I. (2008). Ranking of field association terms using co-word analysis. *Information Processing & Management*, *44*(2), 738–755.

Rousseau, R., Zhang, L., & Hu, X. (2019). Knowledge integration: Its meaning and measurement. In *Springer handbook of science and technology indicators* (pp. 69–94). Cham: Springer.

Shen, Z., Chen, F., Yang, L., & Wu, J. (2019). Node2vec representation for clustering journals and as a possible measure of diversity. *Journal of Data and Information Science*, *4*(2), 81–94.

Sinha, A., Shen, Z., Song, Y., Ma, H., Eide, D., Hsu, B. J., & Wang, K. (2015). An overview of Microsoft academic service (MAS) and applications. In *Proceedings of the 24th international conference on world wide web* (pp. 243–246).

Skupin, A., Biberstine, J. R., & Borner, K. (2013). Visualizing the topical structure of the medical sciences: A self-organizing map approach. *PLoS One*, *8*(3), 16.

Stirling, A. (2007). A general framework for analysing diversity in science, technology and society. *Journal of the Royal Society, Interface*, *4*(15), 707–719.

Wagner, C. S., Roessner, J. D., Bobb, K., Klein, J. T., Boyack, K. W., Keyton, J., ... Börner, K. (2011). Approaches to understanding and measuring interdisciplinary scientific research (IDR): A review of the literature. *Journal of Informetrics*, *5*(1), 14–26.

Waltman, L. (2016). A review of the literature on citation impact indicators. *Journal of Informetrics*, *10*(2), 365–391.

White, H. D. (2003). Author cocitation analysis and Pearson's *r*. *Journal of the American Society for Information Science and Technology*, *54*(13), 1250–1259.

Yan, E., & Ding, Y. (2012). Scholarly network similarities: How bibliographic coupling networks, citation networks, co-citation networks, topical networks, coauthorship networks, and co-word networks related to each other. *Journal of the American Society for Information Science and Technology*, *63*(7), 1313–1326.

Zhang, L., Rousseau, R., & Glanzel, 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.

Zhang, Y., Zhang, G., Chen, H., Porter, A. L., Zhu, D., & Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting and Social Change*, *105*, 179–191.

How to cite this article: Bu Y, Li M, Gu W, Huang W. Topic diversity: A discipline scheme-free diversity measurement for journals. *J Assoc Inf Sci Technol*. 2021;72:523–539. https://doi.org/10.1002/asi.24433

APPENDIX A.: THE LOUVAIN METHOD

The Louvain method aims to optimize the modularity of a network (Newman & Girvan, 2004) for quantifying the closeness of communities in the given partition compared with a null model. In our empirical studies, we first calculate the modularity value, Q, with the formula:

$$Q = \frac{1}{2m} \sum_{V'_\varphi, V'_\omega \in \boldsymbol{V}'} \left(ew_{\varphi\omega} - \frac{k_{V'_\varphi} k_{V'_\omega}}{2m} \right) \delta(\varphi, \omega) \qquad \text{(A1)}$$

In this formula, φ and ω are two nodes in G. m is the number of edges. $weight_{\varphi\omega}$ is the edge weight between nodes φ and ω. k_φ is the degree of φ. $\delta(\varphi, \omega)$ is a binary variable to indicate whether φ and ω belong to the same community ($\delta(\varphi, \omega) = 1$) or not ($\delta(\varphi, \omega) = 0$).

The Louvain method involves two levels of iterations based upon a tree structure. The first level (step B below) uses a bottom-up clustering approach, and the second level (step A below) adopts a swap method that avoids assigning nodes in the same community so they could not be adjusted later. The specific steps of the Louvain method are:

Step A: Initially we regard every node as a community and traverse all nodes. For a certain selected node, by moving it to the community of its neighbor, we can get ΔQ, the variance of Q after moving a node. We try to place the node in the community of every one of its neighbors until we get the maximum ΔQ. If the maximum $\Delta Q \leq 0$, we do not remove this node from its original community; otherwise, this node is placed in the community, leading to the maximum gain of modularity. We keep doing this until Q does not increase anymore.

Step B: We induce nodes in every community (i.e., transform community to node) by implementing step A iteratively until Q does not grow anymore.

Finally, we obtain a function $f: \boldsymbol{V}' \rightarrow \boldsymbol{C}$, where $\boldsymbol{V}'$ is the node set in $\boldsymbol{G}$ and $\boldsymbol{C} = \{1, 2, ..., c\}$ is the community index set. In this way, f attaches each node to a community.

APPENDIX B.: LIST OF 38855 TOPICS

Online available on GitHub (A Microsoft Excel file that contains topic ID (the format is T0-T1-T2-...Tn, while Ti is the community detected by the ith round Louvain Method in the topic detection step), topic size (the number of words in this topic), and hubs (keywords with z-score in the topic higher than 2.5)): https://github.com/PKU-Dragon-Team/Appendix-of-Topic-Diversity.