FACULTEIT ECONOMIE EN BEDRIJFSWETENSCHAPPEN

Management, Strategy and Innovation (MSI)

Comparing different implementations of similarity for disparity measures in studies on interdisciplinarity

Bart Thijs, Ying Huang and Wolfgang Glänzel

MSI_2103

Comparing different implementations of similarity for disparity measures in studies on interdisciplinarity

Bart Thijs[1], Ying Huang[2] and Wolfgang Glänzel[3]

[1] *bart.thijs@kuleuven.be*
KU Leuven, Dept of MSI & ECOOM, 3000 Leuven (Belgium)

[2] *ying.huang@kuleuven.be*
KU Leuven, Dept of MSI & ECOOM, 3000 Leuven (Belgium)

[2] *wolfgang.glanzel@kuleuven.be*
KU Leuven, Dept of MSI & ECOOM, 3000 Leuven (Belgium)

Abstract

The recent literature on interdisciplinarity in quantitative science studies focuses on the development and interpretation of measures of diversity. These measures consider the mutual similarity or distance between disciplines in the applied subject scheme. Most of the studies on this topic are well aware of the importance of the choice of the proper classification scheme as it may lead to large differences in the obtained diversity scores. The central objective of the present study is to investigate the underlying properties of distinct similarity matrices used as starting point in quantifications of disparity. The study screens ten combinations of three different classification schemes and five implementations of citation or reference-based similarities using a version of a cosine similarity. In addition, each of the ten combinations are calculated for nine sliding time windows. The ten combinations are scored on different evaluative criteria both of quantitative and qualitative nature: *stability*, *discriminative power*, *density*, *skewness and deviation* and *ease of calculation*. The study provides the required tools for an informed choice on the appropriate similarity measures in future research on and application of diversity measures. Based on the investigated criteria, the study favors the use of bibliographic coupling on a medium-resolution granularity subject classification.

Introduction

In quantitative science studies of interdisciplinarity, the measurement of diversity takes a prominent place. The concept was proposed by scientometricians in analogy to its application in ecology (e.g., Macarthur, 1965) and still there is a spill-over from the latter field to scientometrics. Thus, the relevant work from ecologists, most notably that by Jost (2006, 2007 and 2009) and Leinster & Cobbold (2012) is often cited in the scientometric literature (e.g. Wang et al., 2015; Zhang et al., 2016), who stated that "*choices of different subject schemes may lead to different diversity results*" and demonstrated this at both the level of individual paper and at the aggregated level of scientific journals. It would be remiss to mention in this context, that the question of the choice of the aggregation level for the measurement of diversity forms part of the subject of a separate study by Huang et al. (2021). This is the reason, why we do deal with this issue in our recent study. Furthermore, we will not focus on the development or the improvement of existing indicators for measurement in the context of IDR, either. The central issue of the present study is laid on the scientometric methods *underlying* the quantification and measurement of diversity in IDR. This issue is twofold; on the one hand, one has to choose the particular (dis-)similarity measure to define the distance (similarity) between individual disciplines and, on the other hand, one has to select the scientometric method for analysing the links leading to the knowledge integrated into the interdisciplinary documents under study. The latter aspect is commonly used in the context of cognitive links that are manifested as citation and/or textual relationships in the published scientific literature. The particular objective – be it document-clustering exercises or network analysis – is in this connection secondary. In the present study we will use bibliographic coupling (BC), co-citation (CC) and direct citation/cross-citation (CRC) and their combinations. But analogously to the

mapping-of-science exercises, also a combination with text-based similarities (cf. Glänzel & Thijs, 2011) are possible to improve the results, most notably in fields, where citations in periodicals play a less significant role as, for instance, in the humanities and in several disciplines in the social sciences.
A further aspect emerges when applying scientometric methods to IDR studies; apart from the already mentioned aggregation level, the issue of granularity emerges as well. This refers to the question of at what subject level interdisciplinarity is to be investigated, a fine-grained topic level, the more moderate sub-field or discipline level or the more general view at the larger research areas. As this raises not only technical questions but also conceptual ones, granularity related issues are discussed in detail in a separate study by Gänzel et al., (2021). Nonetheless, we will extend our analysis to three different levels of granularity within the Leuven-Budapest subject-classification scheme to give evidence of the significance of this issue.
In the study, we will analyse a combination of various methodological settings and using a large dynamic document set retrieved from the three main journal editions of Clarivate Analytics Web of Science Core Collection to find an optimum solution for the implementation of similarity measures in studying the aspects of variety and disparity in interdisciplinarity.

Methodology and Data

Data Source

All documents indexed as articles, letters or reviews in the three journal editions (SCIE, SSCI, A&HCI) of the 2006-2018 volumes of the Web of Science Core Collection (WoS) have been extracted from this database. The data set contains more than 24.4 million so-called citable publications of types 'article, letter or review'. References indexed with the publications are processed and matched with both the cited paper and with similar references in other publications. Each publication is labelled in three classification systems.

Classification Schemes

Three different levels of classification, each with a distinct level of granularity are selected. At the most fine-grained level are the subject categories from the Web of Science Core Collection also known as 'Web of Science Categories'. This classification system holds 255 subject areas, and each journal indexed in this database is assigned to one or more of these subject categories. The system is dynamic as categories can be added yearly (*'Audiology & Speech-Language Pathology'*, *'Nanoscience & Nanotechnology'*) or can become obsolete (*'Biology, Miscellaneous'*). Changes in the system are not in retrospect applied to issues of journals already indexed earlier in the database.
For the analysis at the meso-level, we use the 74 disciplines or subfields of the updated Leuven-Budapest scheme (Glänzel & Schubert, 2003; Glänzel, Thijs & Chi, 2016). The scheme applies a hierarchical system on top of the Web of Science categories. This system is less prone to changes as the total number of disciplines remains stable. Of course, dynamics at the level of individual journals are captured through yearly assignments, allowing shifting profiles over time.
The highest level of aggregation comes from the 16 major subject fields from the same Leuven-Budapest scheme.
The choice of a particular classification scheme is in concordance with the fixed framework assumption stated by Rousseau (2019) and requires that the classification scheme used for final disparity or diversity calculations cannot be altered through the deletion or addition of any particular class.

Similarity Measures

This study proceeds from a vector space model. One of the applicable native similarity measures in a vector space is the uses the cosine measure, which is actually the cosine of the angle between the reference/citation vectors representing the two documents formed by the scalar product of the two vectors divided by the product of their lengths (see equation below). Moreover, in a Boolean vector space as defined by bibliographic coupling and co-citations between individual documents, the cosine measure becomes simply identical with Salton's measure, i.e., the number of joint references/citations divided by the geometric mean of the total number of references/citations of the two documents. Note that this approach applies to individual documents, while cross-citation measures, although methodologically based on direct citations, are applied to document sets and, therefore, require a different approach, particularly for two reasons: The underlying (direct) citation links are not symmetric and unit self-transactions often result in matrices with dominant main diagonals. We will use the formula proposed by Lin et al. (2015) to avoid possible biases caused by this effect (see second equation below).

Based on the choice of underlying citation link a different matrix is subject to this calculation. At each level of classification, a vector is created for the distinct subject classes, and a scalar product is taken between the vectors, here denoted by a and b.

$$similarity(\boldsymbol{a}, \boldsymbol{b}) = \frac{\sum_i^n a_i\, b_i}{\sqrt{\sum_i^n a_i^2}\,\sqrt{\sum_i^n b_i^2}}$$

1. *Bibliographic Coupling (BC):*

   The elements (i) of the vector representing a (sub)field in the vector space are integer values, which express the number of references to documents in the database. In particular, these elements take the values 1 or 0 according as the document i is cited or not. The length of the vector is equal to the number of citable documents in the database.

2. *Indirect Bibliographic Coupling (BC^{ind}):*

   The elements (i) of the vector representing a (sub)field in the vector space are integer values, which express the number of references to all documents assigned to a particular (sub)field in the database. These elements take the values 0, 1, 2, … according to the absolute frequency of the cited references. The length of the vector is equal to the number of (sub)fields in the applied classification scheme. This approach is referred to as *indirect bibliographic coupling* as it links (sub)fields through the field classification of the shared references. This is an indirect link as it allows two (sub)fields that do not refer to the same publication at all to be linked and it introduces a *partial* aggregation, i.e., on the side of the references. The implementation outlined here is similar to that proposed by Leydesdorff and Rafolds (2009) for the creation of global science maps.

3. *Co-Citation (3-year citation window) (CC^{3yr}):*

   The elements (i) of the vector representing a (sub)field in the vector space are integer values, which express the number of citations from documents in the database. Only citing documents indexed in the same year as the cited document or in the two subsequent years are taken into consideration. The elements take the values 1 or 0 according as the document i has cited the document or not in this citation window. The length of the vector is equal to the number of citing documents in the database in the period.

4. *Co-Citation (Open citation window) (CCO):*

   The elements (i) of the vector representing a (sub)field in the vector space are integer values, which express the number of citations from documents in the database. All citing documents, without any restriction with respect to the year of indexing are considered. The length of the vector is equal to the number of citing documents in the database.

5. *Cross-Citation (CRC):*
   While BC and CC are based on individual-document relationships, which can be aggregated to any level, in our study to the subject level of different granularity, cross-citation links, as a combination of references and citations derived from direct citations, are defined on document sets. The elements (i) of the vector representing a document set, i.e., a field or discipline, in the vector space are integer values, which express the number of references and citations to all documents assigned to a particular field in the database. The length of the vector is equal to the number of fields or disciplines in the applied classification scheme. As similarity is usually understood as a symmetric relationship, the resulting matrix needs to be symmetrised. The field or class vector contains binary values that denote whether the field is at the citing or cited side of a particular reference. The length of the vector is equal to the number of references in the database. In fact, the vectors are rows in the incidence matrix describing the undirected and unweighted version of the citation network between the subject fields. This implementation of the cosine similarity results in exactly the same measure as the similarity based on the normalised cross-citation matrix (Zhang et al., 2016) which is as follows

$$s_{ij} = \frac{c_{ij} + c_{ji}}{\sqrt{(TC_i + TC_i)(TC_j + TR_j)}}\ (with\ i \neq j)\ ,$$

   where i and j refer to subject fields, ($c_{ij} + c_{ji}$) denotes to the number of cross-citations between subjects i and j and TC (TR) the total number of citations received by (given to) the subject i and j, respectively from (to) all other subject fields.

Selected combinations

From the 15 possible combinations of granularity levels provided by the Leuven-Budapest classification scheme and the proposed similarity measures, ten have been selected for further analysis. These ten combinations are summarised in Table 1. (Direct) bibliographic coupling is calculated for all three schemes. The two variants of co-citation are applied on the two levels of the Leuven-Budapest scheme as is cross-citation. The indirect version of bibliographic coupling is restricted to the 74 disciplines. As nine publication windows are used for each selected combination, the study is finally based on 90 different datasets. For the complete Leuven-Budapest scheme consult the Appendix and Clarivate Analytics database website (Clarivate Analytics, 2021) for the underlying about 250 WoS categories.

Publication Windows

Nine distinct but overlapping publication windows of three years are used. The first window comprises publications between 2008–2010, the followed by 2009–2011, then 2010–2012, …

until the last one 2016–2018. This scheme provides a kind of 'moving average' approach with slightly smoothing the observed trends. As a result, 90 different similarity matrices are used in this study.

Table 1. Selected combinations of classification schemes and similarity measures

Classification Scheme	BC	BC^{2nd}	CC^{3yr}	CCO	CRC
16 Major Subject Field	*BC_16*		CC^{3yr}_16	*CCO_16*	*CRC_16*
74 Disciplines	*BC_74*	BC^{ind}_74	CC^{3yr}_74	*CCO_74*	*CRC_74*
255 Subject Categories	*BC_255*				

Evaluation Criteria

Before we present the properties of the different combinations, it is reasonable to discuss the different criteria that will be taken into consideration when evaluating the appropriateness, applicability and usability of these implementations of the similarity measures at different levels of granularity. Both quantitative as qualitative criteria are relevant in this context.

Stability

The interpretation of this first criterion is two-fold. We interpret stability primarily in terms of dynamic robustness, that is, as stability of the distribution of the indicator over time. This includes that changes in general descriptive statistics remain within reasonable boundaries. Large changes or fluctuations would influence the final scores that will be obtained. Hence, this time-based instability is to the detriment of the validity of the statement with respect to the dynamics of final scores. This also implies that the same baseline and reference standards may be applied to different periods.

On the other hand, as subject fields are dynamically evolving and so are the relationships between the fields, the underlying similarity matrices calculated over different periods must be able to capture and mirror those in an adequate manner. Stability should not evolve into high rigidity.

Discriminative Power

Also, this criterion is twofold as it indicates the degree to which the similarity is able to assign different scores to pairs of different subjects and, on the other hand and as a matter of course, different scores must be associated with differences in the strength of the relationship between subjects. As such, the discriminative power of a particular implementation of the class similarity also relates to the ability to detect the dynamics in the relations between (sub-)fields and should be balanced with the stability.

Density

The similarity matrix between the different classes at any granularity level can be considered as the weighted adjacency matrix of the network. Non-zero elements indicate the presence of a link between the subjects. At the global level, the density of the network indicates the number of present links compared to the number of possible links. In this study, the matrix representing the network should be as complete as possible, i.e., the number of zero-elements needs to be minimized. Given the dynamics of science, we cannot exclude that non-existent links between two disciplines will never emerge. Interdisciplinarity and new emerging topics are the most striking examples for such changes. And, as the similarity between subjects is often used in the denominator of the calculation of the indicator, setting the value to zero would result in infinite indicator values.

Skewness and deviation

These two statistical functions reflect important properties of empirical distributions in general. As such, they describe the shape and the asymmetry of the distribution of the similarities in our case. Large standard deviations of the similarity distributions combined with low skewness support the discriminative power at the level of individual pairs of disciplines or fields.

Ease of Calculation

Last but not least, the "computability" of the indicators remains an important criterion of applicability and replicability. This criterion particularly refers to the amount of data required to calculate the matrix and to the availability of the data.

Results

The first results with the descriptive statistics for the comparison of the ten selected scenarios and two different time frames are presented in Table 2. Both the mean of the similarity and the density of the network is given.

Table 2. Mean similarity and network density for 10 selected combination given for the first and last 3-year time period considered

	Mean Similarity		**Density of the network**	
	2008-2010	2016-2018	2008-2010	2016-2018
BC_16	0.073	0.100	100%	100%
BC_74	0.024	0.033	97%	99%
BC_255	0.051	0.050	14%	22%
BC^{ind}_74	0.054	0.066	100%	100%
CC^{3yr}_16	0.090	0.098	100%	100%
CC^{3yr}_74	0.067	0.073	100%	99%
CCO_16	0.098	0.098	100%	100%
CCO_74	0.034	0.033	100%	99%
CRC_16	0.042	0.044	100%	100%
CRC_74	0.010	0.009	97%	98%

The first striking observation is that nine out of the ten combinations result in a complete or near-to-complete network. Only the density of the network at the lowest level of granularity with values from 14% to 22% is much lower. This means that most links between subject categories are not present in the selected time windows. As mentioned above, such an absence of pairs of categories in the matrix might pose problems in the calculation of indicator values as it results in undefined indicators. Some of the scenarios do not show any evolution in their mean score. This is the case for the open-ended Co-Citation (CCO), for the Bibliographic Coupling at the level with the finest granularity and for the Cross-Citation. This reflects extreme stability of the obtained similarities but neglects possible structural dynamics that might occur. Breaking down the 16 fields in the ECOOM classification system to the 74 disciplines versions lowers the mean scores. A further breakdown of the 255 subject categories creates multiple edge cuts in the network removing the lowest weights. This explains the higher mean value as non-edges are not set to zero and considered missing values.
In a next step, probability density functions (PDF) are calculated for all ninety distribution. These functions specify the probability that a value or observed similarity is within a particular

range. More formally, this probability would be the result of the integral of that distribution taken over the specific range. Figures 1, 2 and 3 plot these functions for the ten different combinations.

All ten plots confirm the different patterns presented in Table 2. The first four plots in Figure 1 show the PDF for the bibliographic coupling-based combinations. It is immediately clear that the choice of classification scheme has an enormous impact on the obtained similarity measures.

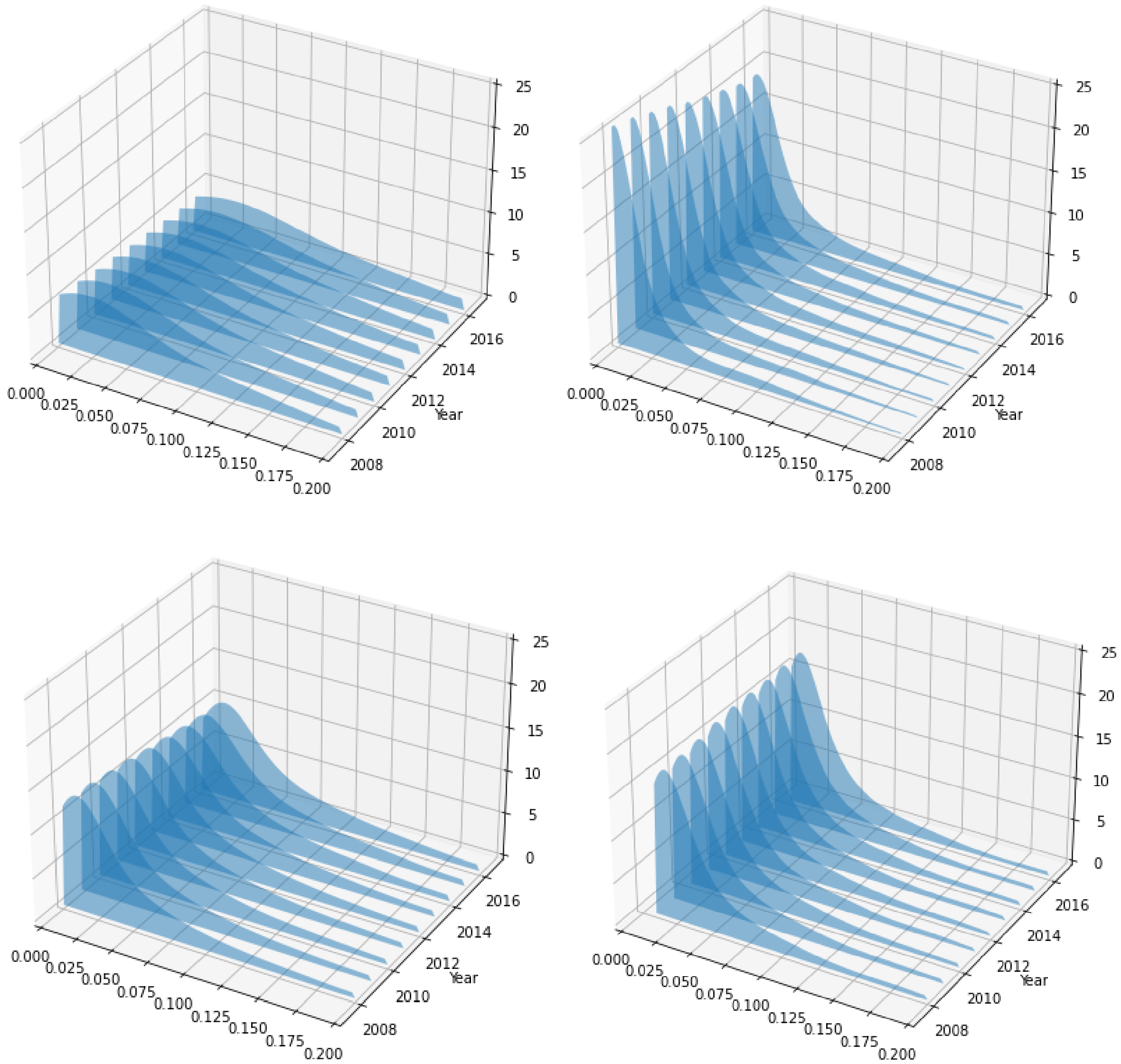

Figure 1. Probability Density Function of the similarities using bibliographic coupling (top left: *BC_16*, top right: *BC_74*, bottom left: BC^{ind}*_74*, bottom right: *BC_255*).

The plot in the top-left corner of Figure 1 uses the 16 fields in the ECOOM classification and has a very flat distribution with high standard deviation and low skewness. This is contrasted by the more skewed distribution of the 74 disciplines alternative in the top-right corner. These plots support the capabilities of this combination to capture structural changes. High-rank correlations between the time slices support claims on the stability of this option.

The indirect version of bibliographic coupling on 74 disciplines, bottom-left in Figure1, shows more variation of the scores. The dynamics are retained but less pronounced, and the mean value only increases by 21% compared to 37% in the (direct) BC. The standard deviation starts

with a higher value and the skewness is lower. This approach, however, suffers from a more complex calculation.
The Bibliographic coupling with the about 250 subject categories has a similar pattern of the probability density functions. The pattern is only calculated for the existing, non-zero values.

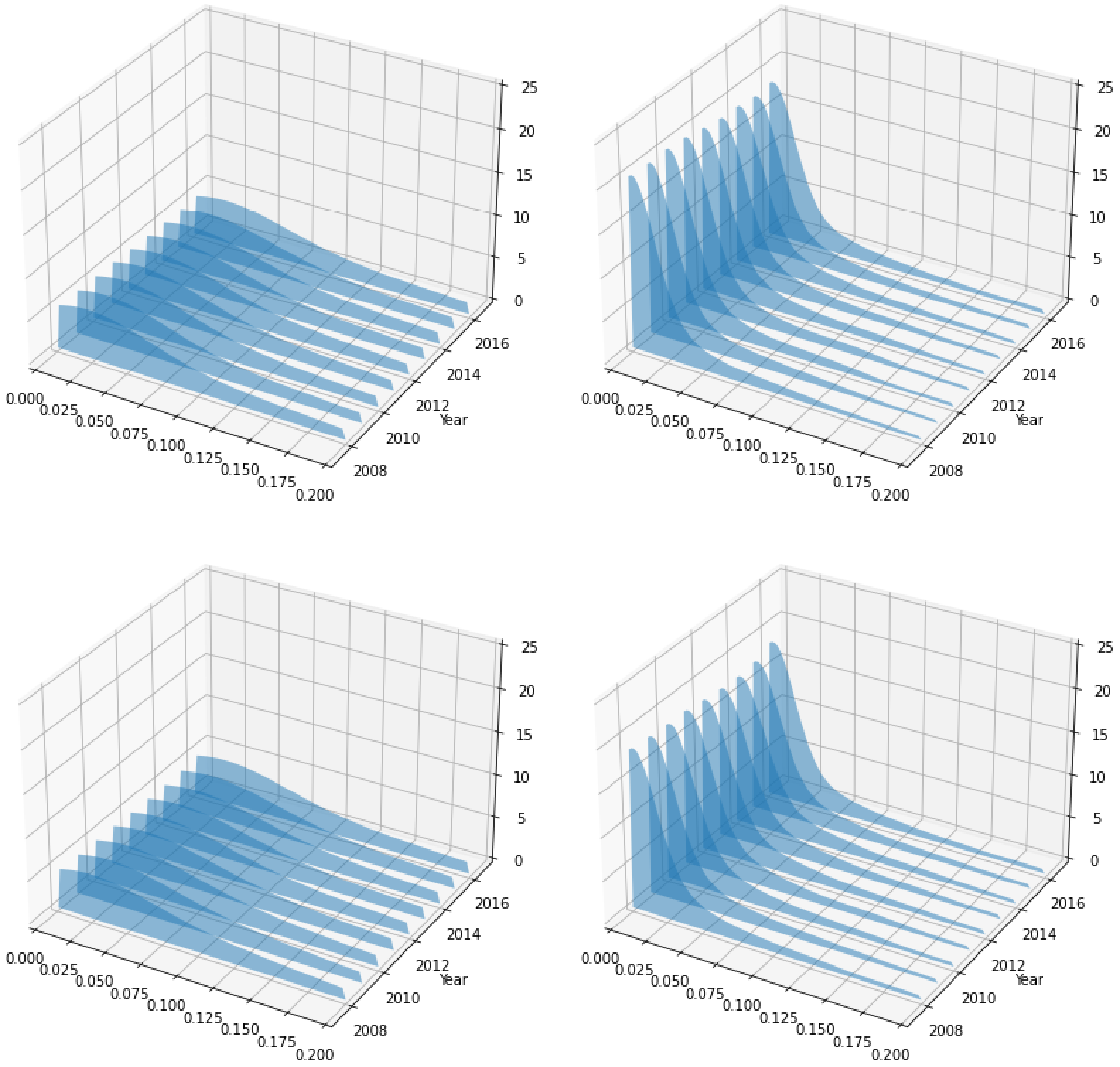

Figure 2. Probability Density Function of the similarities using Co-Citation (top left: *CC^{3yr}_16*, top right: *CC^{3yr}_74*, bottom left: *CCO_16*, bottom right: *CCO_74*.

Figure 2 holds the four alternative versions of the Co-Citation based approach. The plots on the top are based on a three-year citation window, while the bottom row holds the approaches with an open citation window. The differences in flatness between left and right in Figure 2 is analogous to the top row in Figure 1. It is based on the distinction between the 16 and 74 classes in the applied classification scheme. The effect of the applied citation window is marginal. All four plots show a quite rigid pattern of the similarity distributions and underpin the lack of the ability to capture structural changes over time.

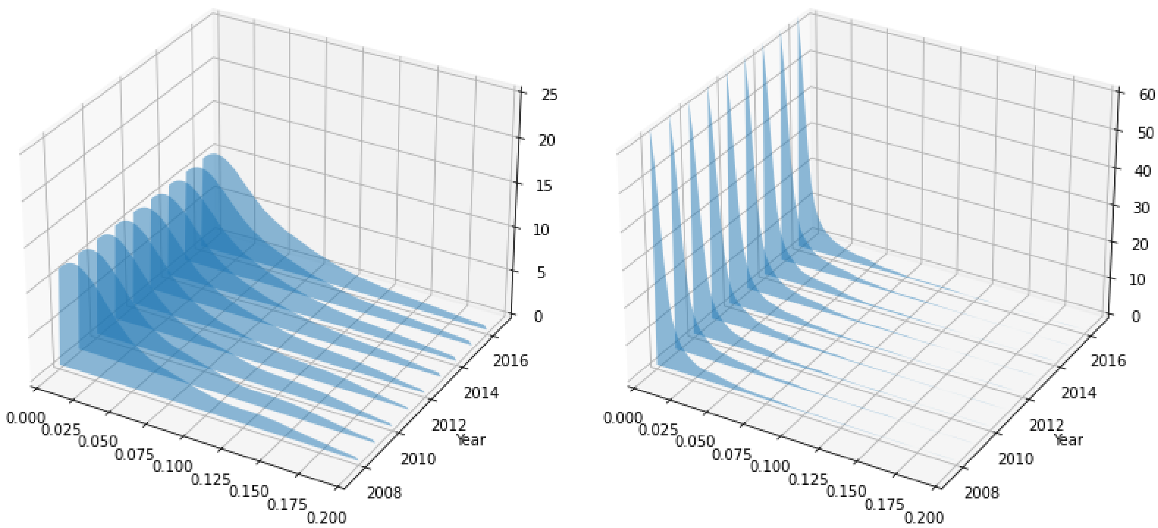

Figure 3. Probability Density Function of the similarities using Cross-Citation (left: *CRC_16*, right: *CRC_74*)

The last two plots are presented in Figure 3. These distributions are more skewed than any of their counterparts with the same underlying classification scheme. In fact, the plot on the right-hand side has an extreme pattern with a steep spike close to zero indicating that almost all obtained similarity scores are very small. This distorts the discriminative power.

Table 3. Summary of the performance of the 10 combinations on five evaluative criteria

	Stability	*Density*	*Discriminative power*	*Skewness & Std Dev*	*Ease of calculation*
BC_16	X	X			X
BC_74	X	X	X	X	X
BC_255	X		X	X	
BC^{ind}_74	X	X	X	X	
CC^{3yr}_16		X			X
CC^{3yr}_74		X	X	X	X
CCO_16		X			X
CCO_74		X	X	X	X
CRC_16		X		X	
CRC_74		X			

Finally, Table 3 provides a summary of the scoring of the ten different scenarios on these five criteria. As mentioned above, these criteria have not only a pure quantitative nature but also require a more qualitative interpretation.

Conclusions

The first observation from the presented analysis relates to the broad range of distributions that are obtained with different combinations of classification schemes and similarity measures. Consequently, this confirms the importance of the proper choice of combination as already raised in the introduction. This choice will have substantial consequences on the final implementation of the disparity and variety measures. Based on the provided material, researchers and users can make the appropriate choice of both classification scheme and similarity measure for their particular application of the diversity score.

Taking all properties and characteristics of each of the combinations into account, this study favors the bibliographic coupling at the level of the 74 disciplines forward as the most appropriate. This combination provides a nearly complete network. It captures the dynamics in the underlying structure and is still relatively easy to calculate.

Acknowledgments

The research underlying this study is funded by the bibliometric part of the multi-year ECOOM project "Interdisciplinary research & Impact" for the Flemish Government (2019–2023). The authors wish to acknowledge this support.

References

Buckland, M. & Gey, F. (1994). The relationship between recall and precision. *Journal of the American Society for Information Science*, 45, 12-19.

Clarivate Analytics (2021), *Web of Science Web Services Expanded. Subject Categories (Ascatype).* Accessible at: http://help.incites.clarivate.com/wosWebServicesExpanded/appendix1Group/ascaCategories.html.

Glänzel, W., & Schubert, A. (2003). A new classification scheme of science fields and subfields designed for scientometric evaluation purposes. *Scientometrics*, 56(3), 357–367.

Glänzel, W., Thijs, B., & Chi, P.S. (2016). The challenges to expand bibliometric studies from periodical literature to monographic literature with a new data source: The book citation index. *Scientometrics*, 109(3), 2165-2179.

Glänzel, W., Thijs, B., & Huang, Y. & (2021), Improving the precision of subject assignment for disparity measurement in studies of interdisciplinary research. FEB Research Report MSI_2104, Report No. MSI_2104.

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

Jost, L. (2007), Partitioning diversity into independent alpha and beta components. *Ecology*, 88(10), 2427-2439.

Jost, L. (2009), Partitioning diversity into independent alpha and beta components. (Correction note on Jost, 2007). *Ecology*, 90(12), 3593-3593.

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

Leydesdorff, L. & Rafols, I. (2009). A Global Map of Science Based on the ISI Subject Categories. *Journal of the American Society for Information Science & Technology*, 60(2), 348-362.

Macarthur, R.H. (1965). Patterns of species diversity. *Biological Reviews*, 40(4), 510-33.

Ochiai, A. (1957). Zoogeographical Studies on the Soleoid Fishes Found in Japan and its Neighbouring Regions-III. *Nippon Suisan Gakkaishi, 22*, 522-525.

Rousseau, R. (2019). On the Leydesdorf-Wagner-Bornmann proposal for diversity measurement. *Journal of Informetrics*. 13, 906-907.

Wang, J., Thijs, B., & Glänzel, W. (2015). Interdisciplinarity and impact: Distinct effects of variety, balance, and disparity. *PloS ONE, 10*(5), e0127298.

Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator of the interdisciplinarity of journals: taking similarity between subject fields into account. *Journal of the association for information science and technology*, 67(5), 1257-1265.

Appendix

The revised Leuven-Budapest classification scheme according to Glänzel et al. (2016)

THE LEUVEN – BUDAPEST CLASSIFICATION SCHEME FOR THE SCIENCES, SOCIAL SCIENCES AND HUMANITIES

0. MULTIDISCIPLINARY SCIENCES
- X0 multidisciplinary sciences

1. AGRICULTURE & ENVIRONMENT
- A1 agricultural science & technology
- A2 plant & soil science & technology
- A3 environmental science & technology
- A4 food & animal science & technology

2. BIOLOGY (ORGANISMIC & SUPRAORGANISMIC LEVEL)
- Z1 animal sciences
- Z2 aquatic sciences
- Z3 microbiology
- Z4 plant sciences
- Z5 pure & applied ecology
- Z6 veterinary sciences

3. BIOSCIENCES (GENERAL, CELLULAR & SUBCELLULAR BIOLOGY; GENETICS)
- B0 multidisciplinary biology
- B1 biochemistry/biophysics/molecular biology
- B2 cell biology
- B3 genetics & developmental biology

4. BIOMEDICAL RESEARCH
- R1 anatomy & pathology
- R2 biomaterials & bioengineering
- R3 experimental/laboratory medicine
- R4 pharmacology & toxicology
- R5 physiology

5. CLINICAL AND EXPERIMENTAL MEDICINE I (GENERAL & INTERNAL MEDICINE)
- I1 cardiovascular & respiratory medicine
- I2 endocrinology & metabolism
- I3 general & internal medicine
- I4 hematology & oncology
- I5 immunology

6. CLINICAL AND EXPERIMENTAL MEDICINE II (NON-INTERNAL MEDICINE SPECIALTIES)
- M1 age & gender related medicine
- M2 dentistry
- M3 dermatology/urogenital system
- M4 ophthalmology/otolaryngology
- M5 paramedicine
- M6 psychiatry & neurology
- M7 radiology & nuclear medicine
- M8 rheumatology/orthopedics
- M9 surgery

7. NEUROSCIENCE & BEHAVIOR
- N1 neurosciences & psychopharmacology
- N2 psychology & behavioral sciences

8. CHEMISTRY
- C0 multidisciplinary chemistry
- C1 analytical, inorganic & nuclear chemistry
- C2 applied chemistry & chemical engineering
- C3 organic & medicinal chemistry
- C4 physical chemistry
- C5 polymer science
- C6 materials science

9. PHYSICS
- P0 multidisciplinary physics
- P1 applied physics
- P2 atomic, molecular & chemical physics
- P3 classical physics
- P4 mathematical & theoretical physics
- P5 particle & nuclear physics
- P6 physics of solids, fluids and plasmas

10. GEOSCIENCES & SPACE SCIENCES
- G1 astronomy & astrophysics
- G2 geosciences & technology
- G3 hydrology/oceanography
- G4 meteorology/atmospheric & aerospace science & technology
- G5 mineralogy & petrology

11. ENGINEERING
- E1 computer science/information technology
- E2 electrical & electronic engineering
- E3 energy & fuels
- E4 general & traditional engineering

12. MATHEMATICS
- H1 applied mathematics
- H2 pure mathematics

13. SOCIAL SCIENCES I (GENERAL, REGIONAL & COMMUNITY ISSUES)
- Y1 education, media & information science
- Y2 sociology & anthropology
- Y3 community & social issues

14. SOCIAL SCIENCES II (ECONOMIC, POLITICAL & LEGAL SCIENCES)
- L1 business, economics, planning
- L2 political science & administration
- L3 law

15. ARTS & HUMANITIES
- K0 multidisciplinary
- K1 arts & design
- K2 architecture
- K3 history & archaeology
- K4 philosophy & religion
- K5 linguistics
- K6 literature

MANAGEMENT, STRATEGY AND INNOVATION (MSI)
Naamsestraat 69 bus 3535
3000 LEUVEN, Belgium
tel. + 32 16 32 67 00
msi@econ.kuleuven.be
https://feb.kuleuven.be/research/MSI/