A Mathematical Approach to Assess Research Diversity: Operationalization and Applicability in Communication Sciences, Political Science, and beyond

Article *in* Scientometrics · September 2020

DOI: 10.1007/s11192-020-03680-6

Scientometrics
https://doi.org/10.1007/s11192-020-03680-6

A mathematical approach to assess research diversity: operationalization and applicability in communication sciences, political science, and beyond

Manuel Goyanes[1,2] · **Márton Demeter**[3] · **Aurea Grané**[4] · **Irene Albarrán-Lozano**[4] · **Homero Gil de Zúñiga**[2,5,6]

Received: 28 March 2020

Abstract
With today's research production and global dissemination, there is growing pressure to assess how academic fields foster diversity. Based on a mathematical problem/solve scheme, the aim of this study is twofold. First, the paper elaborates on how research diversity in scientific fields can be empirically gauged, proposing six working definitions. Second, drawing on these theoretical explanations, we introduce an original methodological protocol for research diversity evaluation. Third, the study puts this mathematical model to an empirical test by comparatively evaluating (1) communication research diversity in 2017, with respect to field's diversity in 1997, and (2) communication research and political science diversity in 2017. Our results indicate that, contrasted to pattern diversity, communication research in 2017 is not a diverse field. However, throughout the years (1997–2017), there is a statistically significant improvement. Finally, the cross-comparison examination between political and communication sciences reveals the latter to be significantly more diverse.

Keywords Research diversity · Diversity · Communication science · Political science · Diversity gaps

In recent decades, research diversity has become a central element in shaping the form and content of scientific fields (Metz et al. 2016), mirroring the growing societal and economic demands and pressures of most democratic societies (Dhanani and Jones 2017). With the growing globalization of academia, diversity enables new opportunities to configure inclusive scientific fields (Waisbord and Mellado 2014; Waisbord 2016), build upon the development of plural approaches to scientific facts and knowledge progress (Stephan and Levin 1991). There is a general consensus that research diversity points to the matureness and

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11192-020-03680-6) contains supplementary material, which is available to authorized users.

✉ Manuel Goyanes
mgoyanes@hum.uc3m.es

Extended author information available on the last page of the article

sophistication of most academic disciplines (Wasserman 2018), enriching empirical evidences with plural visions of the world (Livingstone 2007; Willems 2014), and challenging the taken-for-granted assumptions of academic elites (Demeter 2018). However, despite the importance of rigorously measuring the state in which different intellectual terrains are positioned regarding research diversity, little research has directly developed a reliable instrument to both evaluate diversity claims and infer the potential *diversity gaps* that exist in the academia. This paper seeks to palliate this gap, proposing a protocol to evaluate research diversity, from a multivariate perspective, based on six working definition. We illustrate this protocol in the fields of Communication and Political Science.

For the genesis of this article we took the following approach: initially, we conduct a brief literature review of diversity measures in general, and in bibliometric studies in communication research in particular, with the aim of designing a research diversity framework, conceptualizing the main items and scales often used for gauging research patterns in the field. Despite extant research on communication studies seldom address the potential formulas to measure diversity in research (an exception would be Leydesdorff and Probst 2009), they provide critical perspectives, variables and measurements to assess the evolution of the field (in terms of authorships, methodologies, thematic approaches) and thus the potential diversity of its core components (Freelon 2013; Günther and Domahidi 2017; Walter et al. 2018). After the literature review, we propose, define and describe a methodology and the associated research protocol to calculate the research diversity of a given field and its research production.

Since our interest is in Communication Sciences, we apply these measurements to calibrate this discipline first. Specifically, we conducted a content analysis of a representative and randomized sample of articles ($N = 283$) published in all Journal Citation Reports (JCR) journals ($N_J = 84$) indexed under the category of "communication" in 2017. In addition, we assess the current diversity of research in Communication Sciences compared to that of 20 years ago ($N = 263$; $N_J = 36$), following the same methodological procedure outlined above. Finally, we compare this research diversity with that of a cousin field, i.e. Political Science (N = 329; $N_J = 169$). In all cases, sample sizes were calculated with a confidence level of 95%. Therefore, assuming normality, the final samples had a sampling error of less than 5%.

Measuring diversity: a brief historical overview

Measuring diversity has a long tradition (Rao 1982a). The first attempts to provide reliable diversity measurements date back as those initial efforts of Gini in economics (Gini 1912), Sokal and Sneath in biology (1963), Agresti and Agresti in sociology (1978) or Rao in anthropology (1948). Rao (1982a) reviewed some of these measures and offered three unified approaches for deriving them (Rao 1982a), providing also diversity decomposition examples within a population in terms of given or conceptual factors (Rao 1982b). Later scientometric scholars interested in diversity issues mostly adopt and modify Rao's indices, showcasing the strong influence of Rao's works (Leydesdorff et al. 2019; Stirling 2007), in applying diversity measures on different levels of analysis, including individual journals (Zhang et al. 2009, 2010), and articles (Zhang et al. Zhang et al. 2016).

Stirling (2007), who partially built his approach on Rao's calculations (1982a), considered diversity as an attribute of all systems whose elements could be appointed into different categories. These three systemic features are: variation, balance and disparity.

By reference of ten quality criteria, Stirling proposes a new general diversity heuristic in which each of the aforementioned three subordinate properties—variation, balance and disparity—could be systematically explored. Later scholarships typically adopted Stirling's insights regarding the use of variation, balance and disparity in gauging diversity (Rafols and Meyer 2010; Ráfols 2014).

Bone and his colleagues (Bone et al. 2019) defined diversity in line with Stirling's conceptualization (Stirling 2007), too, but as opposed to Striling (2007) and Ráfols (2014), they measured distances between individuals, and not categories. By conceptualizing diversity on this basis, they followed Boschma work (2005) who established the concept of proximity as a key concept in diversity calibration. Boschma and his later followers applied five forms of proximity, namely cognitive, organizational, social, institutional and geographical proximity, where greater proximity in each category means greater diversity.

More recently, Leydesdorff and Ráfols (2010) analyze different indices by which interdisciplinarity could be quantitatively measured, such as Gini coefficients, Shannon entropy indices, and the Rao-Stirling diversity index. Later research showed that using Rao–Stirling diversity (RS) indices sometimes produces anomalous results (Leydesdorff et al. 2019). It is typically argued that these anomalies could be related to the use of the dual-concept diversity that combines both balance and variety (Stirling 2007). Based on this observation, Leydesdorff et al. (2019) modified RS into an index that operationalizes the three diversity features of Stirling—variety, balance and disparity—independently, and then combines them ex post. This formula has been criticized and slightly modified later by Rousseau (2019).

The contribution of our study is as follows: instead of providing a specific formula or comparing different formulas, we propose an entire protocol to gauge the diversity of a given academic field based on some specific characteristics of its authors and the type and features of the research they carry out. While the Stirling–Rao indices (and also Simpson diversity indices) are measures of the internal diversity of a variable (and the Stirling–Rao index also incorporates a measure of distance between categories), our proposal is based on comparisons to a certain "diversity pattern". For example, in Rafols and Meyer (2010), diversity formulas are used to compare different disciplines through the variable "ref-of-refs" along with a matrix of dissimilarities between disciplines. On the contrary, our concept of "variable diversity" is defined as a battery of measures that allow us to compare the variability of each of the variables of interest with its corresponding pattern. We have illustrated these comparisons using Hellinger's distance, but any other distance function between probability distributions might be valid. Finally, we take the average of all distances as a comprehensive measure of the field variability. We remark that the choice of the distance function is not as important as the calibration of the threshold, from which it will be decided if the variable of interest follows or not the given diversity pattern. This calibration is done via bootstrap.

Communication research patterns: literature review

While we still lack a sound definition for research diversity and a reliable measurement for its calibration, there is a robust body of literature that, either explicitly or implicitly, problematizes diversity issues in communication research. In the following subsections, we present the main empirical contributions of these research branches, explaining how our study contributes to further evaluate diversity claims and infer the position and evolution of single or multiple fields of science.

Methodological, disciplinary and theoretical foundations of diversity in communication studies

Analyses of publication patterns in communication studies can be found as early as 1989, when the special issue *Communication Research* was first published on this topic (Vol 16 Issue 5). In the same year, *Journal of Communication* also dedicated three special issues to analyzing publication patterns, as well as the most frequently assessed subfields in communication research (Vol 43 Issue 3, Vol 54 Issue 4 and Vol 55 Issue 3), showcasing the growing relevance of such meta-scholarship to evaluate the state of the field. Paradoxically, the first citation analysis of communication journals was also published in Paisley (1989), followed by a brand-new research stream on bibliometric or scientometric studies. This study contributes to this research tradition by assessing the empirical, methodological and thematic evolution of the discipline (Funkhouser 1996; Reeves and Borgman 1983; Rice et al. 1988; Borgman 1989; Rogers 1999; Feeley 2008; Bunz 2005; Griffin et al. 2016; Keating et al. 2019).

Extant research on communication research patterns has also addressed issues around its interdisciplinary foundations. For instance, So (1988) found that communication is one of the less diverse fields amongst social sciences, and Smith (2000) also discovered very limited diversity while examining the interdisciplinarity of technical communication journals. Specifically focusing on *Journal of Communication,* Park and Leydesdorff (2009) found there was little citation activity for disciplines other than communication. However, as Zhu and Fu (2019) argue, these studies were limited in many ways:

> Their research scopes were not sufficiently broad enough to reflect the intellectual diversity of the entire field of communication, barely focusing either on shortlisted, top-tier journals (excluding emerging and niche research areas) or on a specific period of time (ignoring the time-evolving nature of the field). The findings mainly offer descriptive information, but not analytical investigations into the possible associations, which thereby confines the research implications (Zhu and Fu 2019, p. 279).

Other scholars investigated specific patterns in communication publication trends. For instance, by analyzing the publication patterns of nine leading journals, Freelon (2013) established the main topics, methods, and citation universes of the field, empirically demonstrating that, in communication research, better-known journals tend to publish work that is quantitative, empirical, epistemologically social-scientific, and American in nature. The major caveat in this spread is that it almost certainly underrepresents work that is "qualitative, purely theoretical, critical, and non-American" (Freelon 2013, p. 22). Thus, what holds for methodological diversity presumably holds for epistemic and thematic diversities, too. Freelon also implemented descriptive statistics to account for such research patterns, complemented with social network analyses. Freelon's findings have been recently extended by Günther and Domahidi (2017), who analyzed the main themes of top-tier journals in communication and found less thematic diversity than expected. Günther and Domahidi (2017) implemented a topic modelling to specify the myriad of topics that articulate communication research, implicitly defining diversity as the distribution of frequencies for each variable under analysis.

Leydersdorff and Probst (2009) considered communication studies as a hybrid research field between political science and social psychology. The authors analyzed cross-citations between journals in all three ISI categories. They found that, with the development of the strength and identity of communication studies as a genuine discipline, the border of communication with social psychology has become sharper than the border with political science.

Besides the analysis of general publication patterns and the interdisciplinary foundations in the field, there is a tradition of scholarship that deals with diversity measures in different segments of the global academy in general, and in communication in particular (Hendrix et al. 2016). Walter et al. (2018) analyzed many aspects of diversity through the examination of articles published in *Journal of Communication* from 1951 to 2016. The study concentrated mostly on diversities in terms of methodology, interdisciplinary perspectives and theoretical foci. Diversity measures thus far were assessed by calculating percentages of different research categories, statistically describing the research tendencies of the field.

More recently, Zhu and Fu (2019) analyzed all the SSCI indexed communication journals with respect to interdisciplinarity. Their study focuses on the longitudinal citation records of communication journals over the past two decades (1997–2016), in order to measure the amount of citations to and from different research fields. Specifically, Zhu and Fu (2019) estimate the diversity of knowledge transfer (including knowledge import and knowledge export) regarding the field of communication. Their method was inspired by network science. Outward citations were measured by out-degree centrality, while inward citations were measured by weighted in-degree centrality. In addition, Zhu and Fu's (2019) study also measured the longitudinal correlation between citation metrics and journal impact factor (JIF), showing that, besides a growing absolute interdisciplinarity, communication scholarship has been faced with stagnant relative interdisciplinarity over the years.

In contrast to former studies, while most typically concentrate on a sole aspect of diversity, like citation patterns (Bunz 2005), interdisciplinarity (Park and Leydesdorrff 2009; Zhu and Fu 2019), or methodological and topical foundations of the field (Freelon 2013; Günther and Domahidi 2017), our study explores and reports multiple variables that account for the holistic vision of the field's diversity. Hence, research diversity is not calibrated as a discrete dimension, but as a complex system made of 15 different variables that extant research has examined separately (Walter et al. 2018). As opposed to former studies that mostly calculate research diversity through descriptive statistics (i.e. frequencies and percentages), our study provides robust mathematical equations and a systematic research protocol aimed at both assessing diversity claims in science and inferring both the evolution and current state of different intellectual terrains.

Gatekeeping and geopolitics: measuring the geographical diversity of editorial boards and authors

The diversity within the editorial boards of communication journals and its relatedness to publication trends and patterns of the field have been also widely studied. Extant research has demonstrated that the discipline is far from being diverse in terms of editorial boards' geographical diversity, and most scholarship has pointed to a significant Western and especially American dominance in this body of governance (Lauf 2005; Demeter 2018; Goyanes 2020). Leeds-Hurwitz (2019) adds that the diversity of editorial boards might correlate with the journals' production model. The author assumes that, at least in communication, open access, especially diamond open access journals, might have a more diverse editorial board than journals under the classic production scheme. Youk and Park's (2019) study examined the geographical diversity and publication patterns of editors and editorial board members in communication journals, showing that the diversity of editorial boards was related to the journal's affiliated association (NCA and ICA), international orientation, and interdisciplinary nature.

The geopolitical diversity of communication journals has also been widely investigated in the last decade (Bunz 2005; Chakravartty et al. 2018; Demeter 2018; Goyanes and Demeter 2020). Ganter and Ortega (2019) argue that, while there is an increasing diversity in communication journals germane to certain Latin-American topics, leading Western journals and conferences are still lacking diversity in terms of Latin-American authors. The geopolitical diversity and intraregional imbalance were measured by descriptive statistics, through which the authors identify, proportionally, the participation of different world regions in the European communications community. Guenther and Joubert (2017), analyzed both gender and geopolitical diversity in science communication journals throughout time, finding that although gender inequalities have decreased slightly, Western dominance remained at a similar level over the years. They measured diversity by analyzing cross-cultural and cross-country collaborations, providing descriptive data on the most productive countries in the field of science communication (i.e. frequencies).

While the aforementioned studies made meaningful contributions towards a better understanding of the long-standing imbalances that exist both in authorship and editorial boards in the field of communications, extant research does not problematize nor provide a robust yardstick to evaluate the field's diversity. As a result, diversity findings are reported in a "diversity vacuum". Additionally, since most studies rely on descriptive statistics (Bunz 2005) or deployed Simpson's diversity indices (Lauf 2005; Demeter 2018), they fail in estimating a benchmark level of diversity to contrast diversity claims in communication studies. Our study provides computable definitions of research diversity and postulates different potential benchmark levels to statistically infer the state and evolution of diversity in academic fields.

Problem statement

This brief recapitulation on how different bibliometric studies have approached diversity in communication hints to the fact that different diversities—in authorship, thematic focus, methodology, interdisciplinarity and so forth—might exist. However, the methodological approaches and research procedures deployed by extant research were mostly based on descriptive statistics of some specific variables, precluding us to delve deeper into the multidimensionality of diversity and establish reliable statistical inferences about the situation and evolution of diversity within and across academic fields. In short, what extant research lacks is a sound yardstick to empirically test diversity claims and infer the potential *diversity gaps* that exist within academia. When does a given scientific field have statistically significant diversity, and how can we establish statistical inferences on its state and evolution? Moreover, how can different scientific fields be statistically measured to yield sound diversity comparisons? This study seeks to address these gaps by providing a mathematically constructed formula with the direct vision to gauge diversity in communication and statistically infer its position germane to a given benchmark population.

Problematizing, defining and measuring research diversity: a protocol

To follow, we present a methodological protocol to measure the research diversity of a given field and the material published (i.e. papers). Although in this study we focus on a representative sample of JCR journals in Communication Sciences, the protocol and the variables measured are both robust and wide enough to transpire onto other scientific fields and units of analysis.

The starting point is a dataset, which is a representative sample of a given population, whose rows are the cases to be evaluated and whose columns are the variables. The protocol to evaluate research diversity is based on four steps:

1. Establish the benchmark: Select the hypothesized marginal probability distributions for all variables. In absence of other information, discrete uniform distribution may be chosen.
2. Select a proper distance function to evaluate the discrepancy between the empirical marginal distribution and the hypothesized. In this work we have chosen Hellinger distance, although other distances (dissimilarities, divergence measures, indexes, etc.) between two probability distributions may be used.
3. Compute *variable diversity* and *field diversity* as explained below.
4. Express any research question of interest as a test of hypothesis and use the proposed statistics based on variable diversity and field diversity to solve the test. To obtain the probability distributions of the test statistics (confidence intervals) implement a row-wise bootstrap in order to preserve the multivariate structure of the data. This may be of importance in case variables are not independent.

In what follows we detail the steps of the protocol. In order to calibrate the position of a given field in terms of research diversity, we must design a benchmark. We labeled this benchmark *diversity pattern,* for which we consider two possible situations: *grounded truth* and *known/given diversity*. First, in the absence of other information, we assume that a grounded truth exists when a given variable has the same proportion or frequencies in each of its values. In terms of Probability Theory, the concept of grounded truth is known as discrete uniform probability distribution (Everitt and Skrondal 2010). For instance, when measuring the gender representation of a given field in terms of first authorship, grounded truth will exist when 50% of the production is authored by male scholars and 50% by female scholars. Second, a known/given diversity will exist when we know the current diversity of a given population, or when we have established it theoretically. For instance, measuring the gender representation of a given field in terms of first authorship, a known/given diversity will exist when (a) we know the frequencies for the gender distribution of a given benchmark population (the world, USA, a continent, the International Communication Association (ICA), all communication scientists, etc.) or (b) when we establish the frequencies for the gender distribution that we theoretically assume to be diverse, for instance 55–45%, 60–40% or 90–10%.

Given that the values were unknown for most of our variables, we took grounded truth as a benchmark and, in the remainder section, we problematize its conceptualization. We assume that grounded truth, if actually exists, is very difficult to concur, since any given journal has its priorities, agendas, expectations and research focus that drive it to employ specific research methodologies, focusing on specialized thematic areas. In addition, according to Knobloch-Westerwick and Glynn (2013), there are gender-oriented topics in Communication Sciences, meaning that some thematic areas are more prone to be built and thus consumed by male or female scholars respectively. Luck might also play a crucial role during the peer-review process, journal selection and data gathering. Geographic imbalances might also have a significant impact on diversity, since as previous studies have demonstrated, Western geographies dominates both research production and editorial boards (Lauf 2005; Demeter 2018), which might suggest that their expectations, agendas and perspectives are crucial to shape communication theory, research and teaching (Curran and Park 2000; Luthra 2015).

Grounded truth serves as an ideal measure, not only to account for the potential impact of luck, but also for the combination of external and internal variables (voluntary or not). These conditions, however, point to potential imbalances and thus the lack of diversity that might exist in the academy. Imbalances in a given field are the product of internal and external forces that struggle for domination and not the result of the selected distribution. However, due to the significant impact that external and internal forces might have in diversity measures, the abovementioned priorities, expectations, orientations, focus, etc. clearly reduce the odds of accounting for a grounded truth. This means that not all values of a given variable hold the same odds in reality, although they potentially have the same odds of being selected. Therefore, the different social and/or organizational agents who discretionally and/or voluntary decide which approach or orientation is worth pursuing in a particular journal are crucial in calibrating diversity and thus mitigating or amplifying the distance from the grounded truth (*diversity gaps*). This voluntary and/or discretional orientation is beyond chance or luck, precluding us to make value judgments and open normative discussions on how a given scientific field should or must be (the contrary would happen with known/given diversity, since the frequencies are known or given). Our results simply point to how distant or close a given variable or field is from its respective ideal, calibrating whether this distance is statistically significant or not. Some variables and fields will arguably be more close to their ideal, suggesting that diversity issues are more socialized. Based on this preliminary problematization, we propose five different definitions for calibrating and comparing research diversity, according to the main objective of the measurement involved. This is translated into the following mathematical terms:

Let $\{X_1, X_2, \dots, X_p\}$ be a set of categorical or discrete variables (available from published papers) and let $c_{i1}, c_{i2} \dots, c_{ik}$ be the different categories or values taken by variable X_i, for $i = 1, \dots, p$. Consider the following definitions:

Grounded truth We say that variable X_i has grounded truth if X_i follows a discrete uniform probability distribution, that is

$$p_i = \left(P\left(X_i = c_{i1}\right), P\left(X_i = c_{i2}\right), \dots, P\left(X_i = c_{ik}\right)\right) = \left(\frac{1}{k}, \frac{1}{k}, \dots, \frac{1}{k}\right) \tag{1}$$

For example, if variable X_i is measuring first author's gender (with $c_{i1} = 1$ for male and $c_{i2} = 2$ for female), grounded truth represents the same probability for males and females to be the first authors of a study in the field of communication. Or, in other, more mathematically precise words, we say that there is grounded truth in gender if Eq. 1 is

$$p_i = \left(P\left(X_i = 1\right), P\left(X_i = 2\right)\right) = \left(\frac{1}{2}, \frac{1}{2}\right).$$

In the case that the diversity pattern is known or given, Eq. 1 becomes

$$p_i = \left(P\left(X_i = c_{i1}\right), P\left(X_i = c_{i2}\right), \dots, P\left(X_i = c_{ik}\right)\right) = \left(p^0_{i1}, p^0_{i2}, \dots, p^0_{ik}\right)$$

with $\sum_{i=1}^{k} p^0_{ij} = 1$.

The set $\{X_1, X_2, \ldots, X_p\}$ serves as *diversity pattern* if each X_i has grounded truth (or follows a known/given diversity distribution), for $i = 1, \ldots, p$. In the case of known/given diversity, note that to establish potential tests, the known/given diversity must be a case, scenario or context, and not the population.

Diversity of a g-group of papers This equation is oriented to calibrate the diversity of a given group of papers with regards to several variables. In particular, the equation estimates how far each variable of interest is from grounded truth. The aim is to compute a distance between the empirical frequencies (calculated from the group of papers) and the theoretical probability (given in Eq. 1), storing all the distances in a vector. Mathematically, the diversity of a group of papers is defined as a vector of distances $(d_{g1}, d_{g2} \ldots, d_{gp})$, where

$$d_{gi} = d\left(f_i, p_i\right), \quad \text{for } i = 1, \ldots, p, \tag{2}$$

$f_i = \left(f_{i1}, \ldots, f_{ik}\right)$ being the vector with the empirical relative frequency distribution of variable X_i in the g-group of papers, p_i the discrete uniform probability distribution (same as before) and d any distance function between discrete probability distributions. Note that one should not compute the empirical relative frequency distribution for only one paper. Thus, the quantity defined in Eq. 2 should be computed for a group n_g of papers ($n_g \geq 10$). Also note that the vector $(d_{g1}, d_{g2} \ldots, d_{gp})$ contains the p distances between the empirical relative frequency distribution of variable $X_i's$ in the g-group and the corresponding discrete uniform probability distribution.

g-group mean diversity This is a scalar measure to summarize the diversity of a given group of papers, taking the mean of the elements of vector $(d_{g1}, d_{g2} \ldots, d_{gp})$. Mathematically, g-group mean diversity is defined as the mean of the p distances $d_{g1}, d_{g2} \ldots, d_{gp}$, that is:

$$\overline{d_g} = \frac{1}{p} \sum_{i=1}^{p} d_{gi} \tag{3}$$

Variable diversity This measure is analogous to the diversity of a g-group of papers, but with the difference that the whole sample of papers is considered, instead of only measuring a small group of them. Variable diversity is defined as the vector of distances $\left(d_{G1}, d_{G2} \ldots, d_{Gp}\right)$, where G is the representative sample of indexed published papers in the research field of interest. In our application, $G = 283$, which is the number of papers that were randomly selected as a representative sample of the Communication Sciences field.

Field diversity This measure is analogous to g-group mean diversity, but computed on the whole sample of papers, computing the mean of the elements of vector $(d_{G1}, d_{G2} \ldots, d_{Gp})$. Field diversity is defined as the mean of the p distances $d_{G1}, d_{G2} \ldots, d_{Gp}$, that is:

$$\overline{d_G} = \frac{1}{p} \sum_{i=1}^{p} d_{Gi} \tag{4}$$

We illustrate the previous concepts and definitions in Figure A1 (see the Online Appendix for detailed information). Following, we describe the protocol to measure the research diversity of a given field. First, scholars interested in applying our diversity measurements need to select a representative and random sample of published papers in the research area of interest, and then a set of variables $\{X_1, X_2, \ldots, X_p\}$ to be measured on each paper. In our application, we have selected a representative, proportional sample of 283 JCR articles in

Communication Sciences and 15 different variables to measure diversity (see the coding book below). Remember that all variables should be categorical or discrete. For each variable X_i, authors need to compute the grounded truth or known/given diversity using Eq. 1. In our case, we compute the grounded truth for all variables, except for first author origin/affiliation and first author gender, for which we assume the true probability distributions given by ICA.

To measure the statistical distance between two probability distributions, authors need to select a statistical function. In our application, we have used the Hellinger distance (Nikulin 1994), which is related to the Bhattacharyya coefficient (Bhattacharyya 1943). Given two discrete probability distributions $P = \left(p_1, p_2, \dots, p_k\right)$ and $Q = \left(q_1, q_2, \dots, q_k\right)$, the Hellinger distance between P and Q is given by

$$d(P, Q) = \sqrt{1 - \sum_{i=1}^{k} \sqrt{p_i q_i}} \tag{5}$$

In the application, we have computed the Hellinger distance between the empirical relative frequency distribution of X_i and the corresponding discrete uniform distribution (in the case of grounded truth), that is, for $i = 3,\dots,p$, we have computed

$$d\left(f_i, p_i\right) = \sqrt{1 - \sum_{j=1}^{k} \sqrt{f_{ij} \frac{1}{k}}}$$

Since we assume a known/given diversity in the case of variables X_1 and X_2, Eq. 5 becomes

$$d\left(f_i, p_i\right) = \sqrt{1 - \sum_{j=1}^{k} \sqrt{f_{ij} p_{ij}^0}}$$

The distance takes values in the [0, 1]-interval, being 0 when variable X_i has the diversity pattern (grounded truth or known/given pattern). Note that distance functions like the Euclidean or Manhattan do not make sense here, since they do not take into account that $\sum_{j=1}^{k} f_{ij} = 1$. This approach can complement other studies, where other metrics, such as Kulback–Leibler divergence, entropy, Simpson's diversity, Rao–Stirling index, among other, are used. After selecting a proper statistical distance function, authors need to compute the variable diversity and store it in a row vector, and then compute the field diversity using Eq. 4. Finally, a bootstrap is needed for the estimation of g-group diversity and g-group mean diversity. First, authors need to bootstrap the representative sample of indexed published papers in order to obtain B groups of n papers, that is, select randomly B groups of n papers. In the application, we have taken $B = 200$ and $n = 10$. It is important to select groups randomly in order to avoid biased estimations.

Using Eq. 2, authors have to compute the g-group diversity, for $g = 1, \dots, B$. Then, they must store each g-group diversity as a row of a $B \times p$ matrix. Call the diversity-matrix to this matrix. Note that each column of the diversity matrix contains the bootstrap distribution of distance d_{gi} $(i = 1, \dots, p)$, that is, the bootstrap distribution of the distance of variable X_i to the grounded truth distribution. Therefore, any summary statistic can be computed on these distributions. We recommend obtaining the corresponding means and medians in order to compare them to the corresponding variable diversity. Finally, using Eq. 3, authors have to compute

g-group mean diversity, $\overline{d_g}$, for each row of the diversity-matrix and, lastly, consider the mean over all the bootstrap samples

$$\overline{d_{dB}} = \frac{1}{B}\sum_{i=1}^{B}\overline{d_g}$$

as an estimation of the g-group mean diversity within the field.

Methodology

We will now describe the methodology and protocol of data gathering and data analysis in detail, which must be followed in order to correctly implement our diversity measurements. First, the interested scholars need to create a pool of research papers from all manuscripts that have been published in a given year. In our case we select 2017, Communication Sciences and the SSCI list of Web of Science (N_J= 84). Then, authors need to make a proportional random sample of the pool of articles that is representative to all research papers published with a margin error of $\pm 5\%$. The random selection can be implemented by using a computerized random number generator. In our case the proportional random sample was $N = 283$.

After the sample selection, independent coders need to content analyze the articles under study. In our case, we follow the Cohen kappa inter-coder agreement coefficient (Cohen, 1960), which adjusts for the proportion of agreements that take place. This was evaluated using the guidelines outlined by Landis and Koch (1977), where the strength of the kappa coefficient is as follows: 0.01–0.20 slight; 0.21–0.40 fair; 0.41–0.60 moderate; 0.61–0.80 substantial; 0.81–1.00 almost perfect. The analysis provided an inter-rater reliability of 97% and a kappa coefficient of 0.93. Therefore, the inter-coder reliability was almost perfect. All discrepancies between coders must be resolved through discussion.

Finally, authors need to create a coding book (see the Online Appendix for detailed information). In order to design and apply the set of diversity measurements previously defined, one must first establish a set of variables which can be oriented to measure the myriad of diversities that might exist in a given field. In our case, we review previous literature on communication research patterns and bibliometric analysis. We consider this stream of research to be crucial in shedding light on diversity issues in Communication Sciences. Although its main purpose is not to calibrate research diversity in the field, it has established reliable measurements to evaluate the evolution of the field, thus indirectly providing relevant variables to shed light on diversity issues (Freelon 2013; Günther and Domahidi 2017; Walter et al. 2018; Demeter 2018, etc.).

All the selected articles were coded manually, since SCI/JCR do not provide data on most of the categories and variables studied. It means that the coders downloaded the randomly selected articles, and manually collected data on authors $\{X_1, X_2, \ldots, X_4\}$ and articles $\{X_5, X_6, \ldots, X_{15}\}$. As a consequence, all the selected articles were content analyzed manually, justifying why it was impossible to conduct "big data" analysis (Gil de Zuniga and Diehl 2017). That is also he main reason to implement a proportional random sample.

Table 1 Variable and field diversity

Category	Variable diversity (distance to the diversity pattern)	% of diversity
X_1	0.1499	85.0
X_2	0.0234	97.7
X_3	0.1810	81.9
X_4	0.4864	51.4
X_5	0.1777	82.2
X_6	0.0740	92.6
X_7	0.1111	88.9
X_8	0.3898	61.0
X_9	0.1530	84.7
X_{10}	0.2224	77.8
X_{11}	0.2748	72.5
X_{12}	0.2849	71.5
X_{13}	0.2681	73.2
X_{14}	0.3142	68.6
X_{15}	0.2078	79.2
Field diversity	0.2212	77.9

Application in the communication sciences field

In Table 1 we give the variable and field diversity according to the set of variables $\{X_1, X_2, \ldots, X_{15}\}$.[1] The interpretation is as follows: grounded truth is 0 (100% diversity) and thus values closer to 0 are more diverse than those farther from 0. As we can observe in Table 1, most variables are close to 0, the first author gender (X_2) being the closest and X_4 (First author affiliation type) the farthest off. The field diversity is 0.2212, i.e. 77.9% diverse.

Regarding variable diversity (see Figure A2 in the Online Appendix for details), we observe that its values are always lower than the median (and the mean) values of the corresponding g-group diversities. Indeed, as the number of papers per group increases, the g-group diversity value gets closer to variable diversity (see Table A1 in the Online Appendix for detailed information).

Our initial analysis indicates some descriptive statistics of research diversity in Communication Sciences in terms of the general field and the variables under study. However, this scrutiny does not provide any empirical evidence regarding the existence of statistically significant differences between grounded truth or the known/given diversity pattern and the field of Communication Sciences (*RQ1*). Similarly, it is important to evaluate possible statistically significant differences between the diversity of each variable under study and grounded truth or the known/given diversity pattern (*RQ2*); and between the field (*RQ3*) and each variable (*RQ4*) diversity in 1997 and grounded truth or the known/given diversity

[1] X_1 = First author affiliation; X_2 = First author gender; X_3 = First author ethnicity; X_4 = First author affiliation type; X_5 = Type of authorship; X_6 = Form of collaboration; X_7 = Interdisciplinarity; X_8 = Area of data collection; X_9 = Methodologies; X_{10} = Research approach; X_{11} = Type of samples; X_{12} = Paradigms; X_{13} = Content area; X_{14} = Analytical focus; X_{15} = Theoretical framework.

pattern. Finally, in order to ascertain how the discipline has evolved over time, the paper also seeks to clarify whether there are statistically significant differences between the field diversity in 1997 and the field diversity in 2017 (*RQ5*); and between each variable diversity in 1997 and each variable diversity in 2017 (*RQ6*), and how each diversity variable ranked according to its contribution in mitigating or amplifying diversity gaps between 1997 and 2017 (*RQ7*).

As a result, we collect data of the same variables under study 20 year ago, following both the methodological procedures and protocols as previously outlined. Therefore, we representatively and randomly select (at 5% margin error) the articles published ($N=263$) in all JCR journals in "communication" in 1997 ($N_J=36$). Based on our research ambitions, and applying the previously defined equations, we aim to answer the following research questions:

Results

RQ1 can be solved by conducting a hypothesis test, to which the null hypothesis is $H_0 : \mu\left(d_G\right) = 0$, where $\mu\left(d_G\right)$ is the expectation (that is, the population mean) of the distance between the field diversity and the diversity pattern (grounded truth or the known/given pattern). Our proposal is to test the null with the following test statistic:

$$\overline{d_G} = \frac{1}{p}\sum_{i=1}^{p} d_{Gi} \tag{6}$$

whose distribution under the null can be obtained by bootstrap. We derived the distribution of the test statistic from $B=20{,}000$ bootstrap samples of size $n=283$ (see Figure A3 in the Online Appendix for a kernel estimation of the density function and Table A2 for the confidence intervals for $\overline{d_G}$).[2] As we may observe, none of them contain the value 0, which means that we should reject the null. However, we must point out that this null hypothesis $H_0 : \mu\left(d_G\right) = 0$ is a very restrictive one, since it implies that there is grounded truth or a known/given diversity pattern in each variable. The explanation is as follows: since a distance cannot take negative values, a sum of distances is equal to zero if, and only if, all the summands are equal to zero. If we look more carefully at the 99%-confidence interval, we can observe that the field diversity is between 0.2133 and 0.2401, meaning that the distance from the diversity pattern (grounded truth or known/given diversity pattern) is between 21.33 and 24.01%, which is not much. Indeed, this confidence interval indicates that, in 2017, the field diversity is between 76 and 78.7%.

To answer *RQ2*, we can conduct p goodness-of-fit tests, with null hypothesis $H_{0i} : \left(f_{i1}, \dots, f_{ik}\right) = \left(p^0_{i1}, \dots, p^0_{ik}\right)$, for $i=1,\ 2$ and $H_{0i} : \left(f_{i1}, \dots, f_{ik}\right) = \left(\frac{1}{k}, \dots, \frac{1}{k}\right)$, for $i=3,\dots,p$. In short, we test if the variables under study follow a known/given probability distribution or a grounded truth (uniform distribution). Therefore, those variables with a p value below $0.05/15=0.0033$ (using Bonferroni correction) are not significant (i.e. are not diverse), while those above 0.0033 are statistically significant and thus diverse. Note that, if a significance level of 0.01 is preferred, then this threshold becomes $0.01/15=0.00{,}067$.

[2] Remind that all bootstrap procedures are done case-wise in order to preserve the multivariate structure of the data, which may be of importance if variables are not independent.

Table 2 Results of the Chi square goodness-of-fit test and 99%-confidence interval

Category	Chi square statistic	*p* value	99%-CI (bootstrap)		% of diversity range	
X_1	85.1651	0.0000	0.1122	0.2154	78	89
X_2	1.2497	0.2636	0.0000	0.0773	92	100
X_3	68.2721	0.0000	0.1288	0.2362	76	87
X_4	526.2721	0.0000	0.4405	0.5475	45	56
X_5	64.1449	0.0000	0.1282	0.2329	77	87
X_6	11.7809	0.0028	0.0239	0.1306	87	98
X_7	26.0106	0.0000	0.0604	0.1676	83	94
X_8	451.8587	0.0000	0.346	0.4698	53	65
X_9	42.2933	0.0000	0.1064	0.2188	78	89
X_{10}	105.7845	0.0000	0.172	0.2808	72	83
X_{11}	138.3004	0.0000	0.2251	0.3319	67	77
X_{12}	198.4629	0.0000	0.2349	0.3409	66	77
X_{13}	182.8445	0.0000	0.2223	0.3269	67	78
X_{14}	293.9293	0.0000	0.2685	0.3688	63	73
X_{15}	224.9894	0.0000	0.1709	0.2607	74	83

In this case, we have conducted the Chi square goodness-of-fit test.[3] Results are shown in Table 2, where we can observe that variable diversity is statistically significant only for First author gender (X_2). Therefore, only this variable follows the diversity pattern, while the others do not. We also show the 99%-confidence intervals obtained by bootstrap. We observe that Form of collaboration (X_6) and Interdisciplinarity (X_7) are not far from grounded truth.

RQ3 can be solved analogously to *RQ1*. Specifically, we are interested in testing $H_0 : \mu\left(d_G^{1997}\right) = 0$, where $\mu\left(d_G^{1997}\right)$ is the expectation (that is, the population mean) of the distance between the field diversity in 1997 and the diversity pattern (grounded truth or the known/given pattern). As before, we use the test statistic of Eq. 6, whose distribution under the null is obtained by bootstrap. We derived the distribution of the test statistic from $B=20{,}000$ bootstrap samples of size $n=263$ (see Figure A4 in the Online Appendix for a kernel estimation of the density function and Table A3 for the confidence intervals for $\overline{d_G}$). Since none of them contain the value 0, we reject the null, meaning that in 1997 the field was not 100% diverse. Indeed, the 99%-confidence interval indicates that the field diversity is between 0.2837 and 0.3197, meaning that the distance from the diversity pattern (grounded truth or known/given diversity) is from 28.37 to 31.975%. Thus in 1997, the field diversity was between 68 and 71.6%, around 7 points lower than in 2017.

RQ4 can be solved analogously to *RQ2*, that is, conducting *p* goodness-of-fit tests, one for each variable. As before, we performed the Chi square goodness-of-fit test.[4] Results are

[3] Note that all expected cell values are greater than 5, hence no Yates correction is needed. For example, if we compute expected cell values in the worst case, which are those corresponding to variable "area of data collection" with k = 13 categories, we have that for a sample size of n = 283, they are n·1/k = 21.77.

[4] Note that, all expected cell values are greater than 5, hence no Yates correction is needed. For example, if we compute expected cell values in the worst case, which are those corresponding to variable "area of data collection" with k = 13 categories, we have that for a sample size of n = 263, they are n·1/k = 20.23.

Table 3 Results of the Chi square goodness-of-fit test and 99%- confidence interval

Category	Chi square statistic	*p* value	99%-CI (bootstrap)		% of diversity range	
X_1	113.7046	0.0000	0.1660	0.2884	71	83
X_2	50.1736	0.0000	0.0991	0.2109	79	90
X_3	172.5057	0.0000	0.2778	0.3858	61	72
X_4	552.3916	0.0000	0.4558	0.5702	43	54
X_5	128.4715	0.0000	0.2508	0.3649	64	75
X_6	97.3308	0.0000	0.1567	0.2687	73	84
X_7	97.2395	0.0000	0.1565	0.2684	73	84
X_8	945.1369	0.0000	0.4668	0.5963	40	53
X_9	73.308	0.0000	0.1488	0.2642	74	85
X_{10}	122.3232	0.0000	0.2462	0.3807	62	75
X_{11}	244.0494	0.0000	0.2683	0.3792	62	73
X_{12}	107.2053	0.0000	0.1950	0.3067	69	81
X_{13}	161.8593	0.0000	0.2643	0.3947	61	74
X_{14}	168.1939	0.0000	0.2954	0.4045	60	70
X_{15}	175.4867	0.0000	0.2399	0.3512	65	76

Sample 1997

shown in Table 3, where we reject the null for all variable diversities at any significance level. Therefore, none of them are 100% diverse. Looking at the 99%-confidence intervals, we can see that First author gender (X_2) is the closest to the diversity pattern.

To solve *RQ5* we have to check if the differences between the field diversity in 1997 and the field diversity in 2017 are statistically significant. Thus, we can perform a test with null hypothesis $H_0 : \mu\left(d_G^{1997}\right) = \mu\left(d_G^{2017}\right)$. Our proposal is to test the null with the following test statistic:

$$\text{Diff\_field} = \overline{d_G^{1997}} - \overline{d_G^{2017}}$$

whose support is the interval [− 1, 1]. The distribution of the statistic under the null is computed from $B = 20{,}000$ bootstrap samples of sizes $n_1 = 263$ and $n_2 = 283$ (see Figure A5 in the Online Appendix for a kernel estimation of the density function and Table A4 for the confidence intervals). We can observe that both limits are positive, indicating that the field diversity in 2017 is closer to the diversity pattern (grounded truth or known/given pattern) than in 1997. Since both limits are positive (zero is not included in the confidence interval), we conclude that there are statistically significant differences between the field diversity in 1997 and 2017.

To answer *RQ6*, we proceed analogously to *RQ5* and obtain 15 bootstrap confidence intervals in order to test if there are statistically significant differences between each variable diversity in 1997 and 2017. We propose the following tests statistics:

$$\text{Diff\_variable}(i) = d_{Gi}^{1997} - d_{Gi}^{2017}, \quad \textit{for } i = 1, \ldots, 15.$$

with support in [− 1, 1]. Their distributions are obtained from $B = 20{,}000$ bootstrap samples of sizes $n_1 = 263$ and $n_2 = 283$. Table 4 contains the confidence intervals (see Figure A6

Table 4 Confidence intervals for Diff_variable statistic

Category	99%-CI (bootstrap)		
X_1	−0.0185	0.1406	
X_2	0.0543	0.1967	***
X_3	0.0744	0.2297	***
X_4	−0.0639	0.0988	
X_5	0.0470	0.2047	***
X_6	0.0571	0.2109	***
X_7	0.0220	0.1744	***
X_8	0.0368	0.2139	***
X_9	−0.0356	0.1236	
X_{10}	−0.0015	0.1584	
X_{11}	−0.0314	0.1205	
X_{12}	−0.1143	0.0427	
X_{13}	−0.0298	0.1336	
X_{14}	−0.0448	0.1111	
X_{15}	0.0073	0.1513	***

***Stands for statistically significant

in the Online Appendix for the kernel density estimations). If we look at 99%-confidence intervals, we can see that the variables which show statistically significant differences in their diversity between 1997 and 2017 are: First author gender (X_2), First author ethnicity (X_3), Type of authorship (X_5), Form of collaboration (X_6), Interdisciplinarity (X_7), Land of data collection (X_8) and Theoretical framework (X_{15}). All of them have experienced a significant diversity increase within these 20 years.

As we can observe in Fig. 1 (*RQ7*) there is a notable increase in the percentage of diversity between 1997 and 2017 in the vast majority of variables. Only for research paradigm (X_{12}) the percentage of diversity in 1997 was greater than in 2017, but this difference was not statistically significant. Therefore, as the field becomes more mature, the diversity gaps are generally mitigated, in most cases significantly, while the diversity gap in 1997 is higher than in 2017 in only one case.

Theoretical applications and more empirical testing: cross-comparisons between academic fields

The application of our diversity measurements can also be implemented to calibrate, compare and rank academic fields. The different variables under study can be adapted or complemented with other values, as long as the studied category (i.e. variable) remains the same across all disciplines. For instance, X_1 (First author origin/affiliation), X_2 (First author gender), X_3 (First author ethnicity), X_4 (First author affiliation type), X_5 (Type of authorship), X_6 (Form of collaboration), X_7 (Interdisciplinary) and X_8 (Land of data collection) are variables whose values should not change much across the spectrums of both natural and social sciences. However, X_9 (Methodologies), X_{10} (Research approach), X_{11} (Type of samples), X_{12} (Paradigms), X_{13} (Content area), X_{14} (Analytical focus), X_{15} (Theoretical framework) are variables with values that should be adapted and/or complemented to capture the nature of each field under study. Nevertheless, in order to make sound

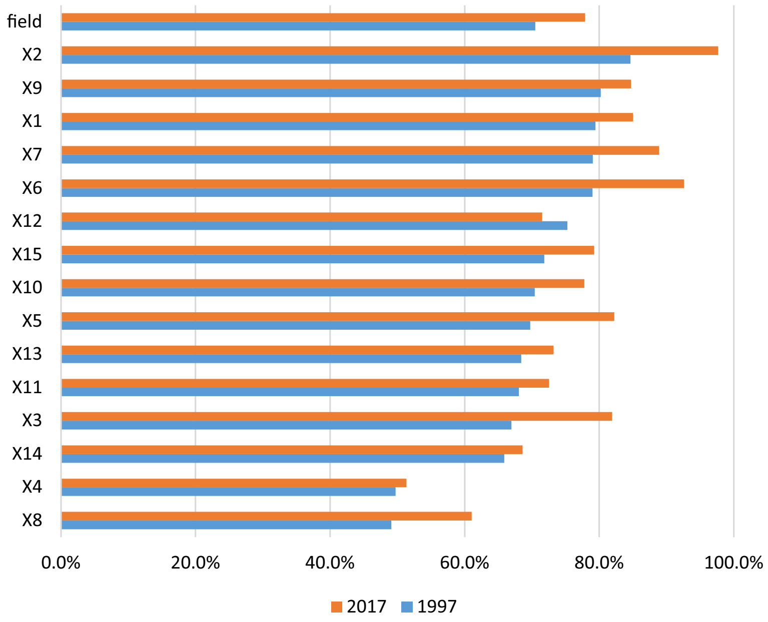

Fig. 1 Diversity gaps between variables in 1997 and 2017

comparisons, every variable under analysis should be added in all fields, modifying or maintaining the values for its measurement. Therefore, when comparing academic fields, variables must remain the same across the board, while values can be adapted, modified, changed or complemented.

The previous application of different diversity measurements was based on a single academic field, i.e. communication sciences, comparing two different points in time (current situation vs. 20 years ago). However, in this section, we apply said diversity measurement to calibrate the diversity distance between two academic fields: communication and political science. First, from a statistical point of view, different academic fields (i.e. academic field "A" and academic field "B") can be considered similar to an academic field in a particular year (i.e. 1997 or 2017). Therefore, this new scenario can be solved following previous indications, in particular those from *RQ5* to *RQ7*. Indeed, when comparing two different academic fields, we are interested in testing $H_0 : \mu\left(d_G^A\right) = \mu\left(d_G^B\right)$, thus we use the test statistic previously proposed:

$$\text{Diff\_field} = \overline{d_G^A} - \overline{d_G^B}$$

Second, to test if there are statistically significant differences between each variable diversity in Field A and Field B, the following test statistics are proposed:

$$\text{Diff\_variable}(i) = d_{Gi}^A - d_{Gi}^B, \quad for\ i = 1, \dots, 15.$$

In consequence, we compare these two different fields. Concerning paper selection for Political Sciences, we chose the same analogous method that we used for communication, leading to a proportional random sample of $N=329$ papers (inter-rater reliability of 95%

Table 5 Confidence intervals for Diff_variable statistic for the comparison between two academic fields

Category	99%-CI (bootstrap)		
X_1	0.0464	0.1918	***
X_2	−0.0645	0.0523	
X_3	0.0412	0.1874	***
X_4	−0.1269	0.0265	
X_5	−0.0567	0.0885	
X_6	0.0587	0.2036	***
X_7	0.0737	0.2204	***
X_8	−0.1558	0.0157	
X_9	0.0114	0.1619	***
X_{10}	−0.1449	0.0020	
X_{11}	−0.0153	0.1340	
X_{12}	−0.1857	−0.0387	***
X_{13}	−0.1796	−0.0322	***
X_{14}	−0.1381	0.0038	
X_{15}	0.1004	0.2369	***

***Stands for statistically significant

and a kappa coefficient of 0.90). Regarding the diversity pattern, we compute the grounded truth for all variables, except for first author origin/affiliation and first author gender, for which we assume the true probability distributions given by IPSA (International Political Science Association).

The distributions of the previous statistics under the null are computed from 20,000 bootstrap samples of sizes $n_1=329$ and $n_2=283$ (see Figure A7 in the Online Appendix for a kernel estimation of the density function and Table A5 for the corresponding confidence intervals). We can observe that both limits are positive, meaning that the field diversity for Communication Sciences is closer to the diversity pattern (grounded truth or known/given pattern) than for Political Sciences. Since both limits are positive (zero is not included in the confidence interval), we conclude that there are statistically significant differences between both academic fields.

Concerning variable diversity, Table 5 contains the corresponding confidence intervals (see Figure A8 in the Online Appendix for the kernel density estimations). If we look at 99%-confidence intervals, we can see that the variables which show statistically significant differences in their diversity between both fields are: First author origin/affiliation (X_1), First author ethnicity (X_3), Form of collaboration (X_6), Interdisciplinarity (X_7), Methodologies (X_9), Paradigms (X_{12}), Content area (X_{13}) and Theoretical framework (X_{15}). In particular, Communication has more diversity than Political Sciences in First author origin/affiliation, First author ethnicity, Form of collaboration, Interdisciplinarity, Methodologies and Theoretical Framework; whereas the contrary occurs in the Paradigms and Content area.

Finally, as we can observe in Fig. 2, the diversity in Communication is greater than that of Political Science in eight out of fifteen variables, although those differences were statistically significant in only six of them.

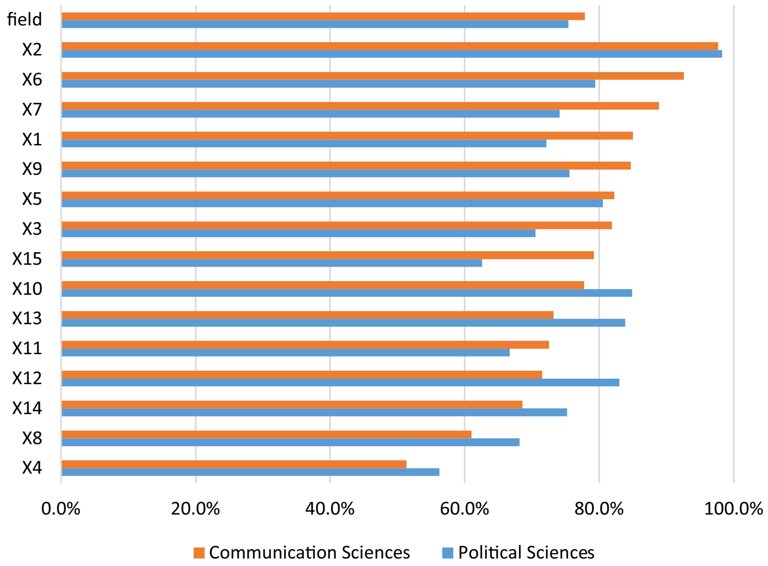

Fig. 2 Diversity gaps between variables in political sciences and communication sciences

Discussion and conclusion

The goal of this study was to propose and test a methodological protocol to calibrate the research diversity in a given scientific field. Specifically, we tested the mathematical feasibility of our instrument within the fields of Communication and Political Sciences. This study offers three inter-related contributions regarding this line of inquiry at different levels of analysis: theoretical, methodological and empirical. First, we propose six theoretical definitions to empirically measure research diversity, describing their mathematical and theoretical foundations in detail: grounded truth, known/given diversity, diversity of a g-group of papers, g-group mean diversity, variable diversity and field diversity. While extant research in ecology (Simpson's Index by Magurran 1988), economics (Hirschmann-index by Hirschmann 2018) and information sciences (Shannon index by Shannon 1948) have provided different equations that may be applied to assess diversity in a myriad of realms, our contribution extends these indices by designing ad hoc measurements to empirically calibrate the potential and multiple dimensions of diversity in science. The 15 categories proposed are thus aimed to capture a detailed portrait of the field diversity in Communication, also adding a temporal frame for longitudinal examination.

Second, we present and describe a research protocol for a step-by-step evaluation of how the different measurements should be applied following standard procedures of data collection and analysis. After proposing a research protocol and problematizing the potential adaptation of our instrument to calibrate diversity in different academic fields, we empirically apply it to evaluate the state of communication, comparing the diversity state in 2017 with the situation twenty years ago. Our empirical evidences demonstrated that diversity should be calibrated as a complex phenomenon and thus different dimensions must be considered. As a result, a given field may hold almost grounded truth diversity in one category, while still lacking it in other variables, as our results demonstrate. In addition, as contrasted with former cross-sectional research (Lauf 2005; Demeter 2018), a longitudinal analysis adds a better understanding of the phenomena, addressing how different features of research diversity may evolve during the course of the years, also signaling potential diversity gaps that may exist in a given field.

In our analysis of the Communication Sciences field, we show that, comparing it to grounded truth or given/known diversity, most variables and the field as a whole are not statistically significant (i.e. are not diverse), suggesting that the discipline still has room for improvement at its macro and micro levels of inclusiveness. In this regard, only the variable "first author gender" is statistically significant, demonstrating that the knowledge production of communication research, taking the ICA as baseline, is representative of its members. The longitudinal analysis also shows that the field is improving its overall levels of diversity throughout the years, as the research production in 2017 has a statistically significant increase in diversity compared to that of 1997. Our results thus suggest that most scientific stakeholders aim to create a more open space for communication research, in which different diversity dimensions may harmoniously coexist. Finally, in order to account empirically for cross-comparisons between scientific fields, our analysis applies diversity measures to calibrate the diversity distance between two cousin fields: Communication and Political Sciences. Our findings show that Communication, compared to Political Sciences, is a significantly more diverse field, especially in terms of first author origin, ethnicity, interdisciplinarity and the methods employed.

In summary, the main purpose of our study is to systematize a general and generalizable protocol for measuring diversity within different academic fields. Therefore, our main ambition is to define a protocol that measures the diversity of a discipline in a multivariate way, based on the information on their authors and the type and characteristics of the research they carry out. Specifically, we measure the diversity of a discipline through the analysis of a multivariate sample of articles published in JCR. For each of the variables of interest, the distance to a reference standard or, in its absence, to the discrete uniform distribution (since we consider that a variable is more diverse the more balanced its probability distribution) is calculated. Our protocol is assumed to be general enough to be applicable to other disciplines.

This study has some limitations that should be addressed by future research. First, while we aimed to be consistent with the categorization schema of former studies (Lauf 2005; Demeter 2018), the geographical coding could be different, nuancing the final results. Second, and most importantly, in order to establish our benchmark comparisons, we rely on grounded truth when frequency distributions were unknown and on given/known diversity when such data was potentially available (in our case from ICA or IPSA for gender and geographical diversity). While our measurements work well and provide sound results for comparisons (between years and across fields), as the benchmark is always the same (although it is not perfect) for gauging the diversity of a given field in a given point of time (i.e. 2017), results may change according to the benchmark of selection. A potential solution to establish a more reliable benchmark for given/known diversity when studying scientific fields in a given point of time is to content analyze a more open scientific ranking (Scopus) and then adjust the frequencies for each variable to the data gathered from JCR journals.

Raising the level of diversity in the global academy in general, and in communication studies in particular, has been a topic of emerging interest in the last decades. The discussions concerning the internalization and diversification of the field are rife with both empirical analyses (Lauf 2005; Demeter 2018; Toth 2018) and theoretical polemics (Waisbord and Mellado 2014; Waisbord 2019), while an inferential examination of research diversity in Communication Studies has been missing. This article contributes to current discussions on research diversity by providing a mathematical apparatus and research protocol for diversity calibration, accounting for the inherent complexity and multidimensionality of the phenomenon and its potential adaptation to other fields. The mathematical definitions proposed could be of great interest for all academics and policymakers oriented to grasp the complexity and evolution of diversity in science, and all those stakeholders who want to establish a more inclusive and diverse global science.

References

Agresti, A., & Agresti, B. F. (1978). Statistical analysis of qualitative variation. In K. F. Schussler (Ed.), *Social methodology* (Vol. 9, pp. 204–237). New York: Wiley.

Bhattacharyya, A. (1943). On a measure of divergence between two statistical populations defined by their probability distributions. *Bulletin of the Calcutta Mathematical Society, 35,* 99–109.

Bone, F., Hopkins, M. M., Ráfols, I., Molas-Gallart, J., Tang, P., Davey, G., & Carr, A. M. (2019). DARE to be different? Applying diversity indicators to the evaluation of collaborative research projects. *Science Policy Research Unit—SPRU working paper series 2019–09*, University of Sussex, UK.

Borgman, C. L. (1989). Bibliometrics and scholarly communication: Editor's introduction. *Communication Research, 16*(5), 583–599.

Boschma, R. (2005). Proximity and innovation: A critical assessment. *Regional Studies, 39*(1), 61–74.

Bunz, U. (2005). Publish or perish: A limited author analysis of ICA and NCA journals. *Journal of Communication, 55*(4), 703–720. https://doi.org/10.1111/j.1460-2466.2005.tb03018.x.

Chakravartty, P., Kuo, R., Grubbs, V., & McIlwain, C. (2018). #CommunicationSoWhite. *Journal of Communication, 68*(2), 254–266. https://doi.org/10.1093/joc/jqy003.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement, 20*(1), 37–46.

Curran, J., & Park, M. (2000). *De-Westernizing media studies*. London: Routledge.

Demeter, M. (2018). Changing center and stagnant periphery in communication and media studies: National diversity of major international journals in the field of communication from 2013 to 2017. *International Journal of Communication*, *12*, 29.

Dhanani, A., & Jones, M. J. (2017). Editorial boards of accounting journals: Gender diversity and internationalisation. *Accounting, Auditing & Accountability Journal, 30*(5), 1008–1040. https://doi.org/10.1108/AAAJ-08-2014-1785.

Everitt, B. S., & Skrondal, A. (2010). *The Cambridge dictionary of statistics* (4th ed.). New York: Cambridge University Press.

Feeley, T. H. (2008). A bibliometric analysis of communication journals from 2002 to 2005. *Human Communication Researh, 34*(3), 505–520. https://doi.org/10.1111/j.1468-2958.2008.00330.x.

Freelon, D. (2013). Co-citation map of 9 comm journals, 2003–2013. Retrieved May 5, 2020, from http://dfreelon.org/2013/09/05/co-citation-map-of-9-comm-journals-2003-2013/.

Funkhouser, E. T. (1996). The evaluative use of citation analysis for communication journals. *Human Communication Research, 22*(4), 563–574. https://doi.org/10.1111/j.1468-2958.1996.tb00379.x.

Ganter, S. A., & Ortega, F. (2019). The invisibility of Latin American Scholarship in European media and communication studies: Challenges and opportunities of de-westernization and academic cosmopolitanism. *International Journal of Communication, 13,* 68–91.

Gil de Zuniga, H., & Diehl, T. (2017). Citizenship, social media, and big data: Current and future research in the social sciences. *Social Science Computer Review, 35*(1), 3–9.

Gini, C. (1912). *Variabiliti e Mutabiliti. Studi Economicoaguridici della facotta di Giurisprudenza dell*. Cagliari: Universite di Cagliari III, Parte II.

Goyanes, M. (2020). Editorial boards in communication sciences journals: Plurality or standardization? *International Communication Gazette*, *82*(4), 342–364.

Goyanes, M., & Demeter, M. (2020). How the geographic diversity of editorial boards affects what is published in JCR-ranked communication journals. *Journalism & Mass Communication Quarterly*. https://doi.org/10.1177/1077699020904169.

Griffin, D. J., Bolkan, S., Holmgren, J. L., & Tutzauer, F. (2016). Central journals and authors in communication using a publication network. *Scientometrics, 106*(1), 91–104. https://doi.org/10.1007/s11192-015-1774-4.

Guenther, L., & Joubert, M. (2017). Science communication as a field of research: Identifying trends, challenges and gaps by analysing research papers. *Journal of Science Communication, 16*(2), 1–19. https://doi.org/10.22323/2.16020202.

Günther, E., & Domahidi, E. (2017). What communication scholars write about: An analysis of 80 years of research in high-impact journals. *International Journal of Communication, 11,* 3051–3071.

Hendrix, K. G., Mazer, J. P., & Hess, J. A. (2016). Forum: Diversity and scholarship on instructional communication. *Communication Education, 65*(1), 105–127.

Hirschman, A. O. (2018). *National power and the structure of foreign trade*. Berkeley: University of California Press.

Keating, D. M., Richards, A. S., Palomares, N. A., Banas, J. A., Joyce, N., & Rains, S. A. (2019). Titling practices and their implications in communication research 1970–2010: Cutesy cues carry citation consequences. *Communication Research*. https://doi.org/10.1177/0093650219887025.

Knobloch-Westerwick, S., & Glynn, C. J. (2013). The Matilda effect—Role congruity effects on scholarly communication: A citation analysis of communication research and journal of communication articles. *Communication Research, 40*(1), 3–26. https://doi.org/10.1177/0093650211418339.

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159–174.

Lauf, E. (2005). National diversity of major international journals in the field of communication. *Journal of Communication, 55*(1), 139–151. https://doi.org/10.1111/j.1460-2466.2005.tb02663.x.

Leeds-Hurwitz, W. (2019). Moving (slowly) toward understanding knowledge as a global commons. *Journal of Multicultural Discourses*. https://doi.org/10.1080/17447143.2019.1695806.

Leydesdorff, L., & Probst, C. (2009). The delineation of an interdisciplinary specialty in terms of a journal set: The case of communication studies. *Journal of the American Society for Information Science and Technology, 60*(8), 1709–1718.

Leydesdorff, L., & Rafols, I. (2010). Indicators of the interdisciplinarity of journals: Diversity, centrality, and citations. *Journal of Informetrics, 5*(1), 87–100.

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.

Livingstone, S. (2007). Internationalizing media and communication studies: Reflections on the International Communication Association. *Global Media and Communication, 3*(3), 273–288.

Luthra, R. (2015). Transforming global communication research with a view to the margins. *Communication Research and Practice, 1*(3), 251–257. https://doi.org/10.1080/22041451.2015.1079156.

Magurran, A. E. (1988). *Ecological diversity and its measurement*. Princeton, NJ: Princeton University Press.

Metz, I., Harzing, A. W., & Zyphur, M. J. (2016). Of journal editors and editorial boards: who are the trailblazers in increasing editorial board gender equality? *British Journal of Management, 27*(4), 712–726. https://doi.org/10.1111/1467-8551.12133.

Nikulin, M. S. (1994). Hellinger distance. In *Encyclopedia of mathematics*. Retrieved May 5, 2020, from https://www.encyclopediaofmath.org/index.php/Hellinger_distance.

Paisley, W. (1989). Bibliometrics, scholarly communication, and communication research. *Communication Research, 16*(5), 701–717. https://doi.org/10.1177/009365089016005010.

Park, H., & Leydesdorff, L. (2009). Knowledge linkage structures in communication studies using citation analysis among communication journals. *Scientometrics, 81*(1), 157–175. https://doi.org/10.1007/s11192-009-2119-y.

Ràfols, I. (2014). Knowledge integration and diffusion: Measures and mapping of diversity and coherence. In Y. Ding, R. Rousseau, & D. Wolfram (Eds.), *Measuring scholarly impact* (pp. 169–190). Cham: Springer.

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

Rao, C. R. (1948). The utilization of multiple measurements in problems of biological classification. *Journal of the Royal Statistical Society, B, 13,* 159–193.

Rao, C. R. (1982a). Diversity and dissimilarity coefficients: A unified approach. *Theoretical Population Biology, 21*(1), 24–43.

Rao, C. R. (1982b). Diversity: Its measurement, decomposition, apportionment and analysis. *Sankhya: The Indian Journal of Statistics, Series A, 44*(1), 1–22.

Reeves, B., & Borgman, C. L. (1983). A bibliometric evaluation of core journals in communication research. *Human Communication Research, 10*(1), 119–136. https://doi.org/10.1111/j.1468-2958.1983.tb00007.x.

Rice, R. E., Borgman, C. L., & Reeves, B. (1988). Citation networks of communication journals, 1977–1985 cliques and positions, citations made and citations received. *Human Communication Research, 15*(2), 256–283. https://doi.org/10.1111/j.1468-2958.1988.tb00184.x.

Rogers, E. M. (1999). Anatomy of the two subdisciplines of communication study. *Human Communication Research, 25*(4), 618–631. https://doi.org/10.1111/j.1468-2958.1999.tb00465.x.

Rousseau, R. (2019). Correspondence. On the Leydesdorff–Wagner–Bornmann proposal for diversity measurements. *Journal of Informetrics, 13,* 906–907.

Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*(3), 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x.

Smith, E. O. (2000). Strength in the technical communication journals and diversity in the serials cited. *Journal of Business and Technical Communication, 14*(2), 131–184. https://doi.org/10.1177/105065190001400201.

So, C. Y. (1988). Citation patterns of core communication journals: An assessment of the developmental status of communication. *Human Communication Research, 15*(2), 236–255. https://doi.org/10.1111/j.1468-2958.1988.tb00183.x.

Sokal, R. R., & Sneath, P. H. A. (1963). *Principles of numerical taxonomy*. San Francisco: Freeman.

Stephan, P. E., & Levin, S. G. (1991). Inequality in scientific performance: Adjustment for attribution and journal impact. *Social Studies of Science, 21*(2), 351–368. https://doi.org/10.1177/030631291021002007.

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

Toth, J. (2018). "U.S. journals can afford to remain regional, but we can not." Author distribution-based internationality of Eastern European communication journals. *KOME—An International Journal of Pure Communication Inquiry, 6*(2), 1–15. https://doi.org/10.17646/KOME.2018.21.

Waisbord, S. (2016). Communication studies without frontiers? Translation and cosmopolitanism across academic cultures. *International Journal of Communication, 10*(2016), 868–886.

Waisbord, S. (2019). *Communication. A post-discipline*. London: Polity Press.

Waisbord, S., & Mellado, C. (2014). De-westernizing communication studies: A reassessment. *Communication Theory, 24*(4), 361–372. https://doi.org/10.1111/comt.12044.

Walter, N., Cody, M. J., & Ball-Rokeach, S. J. (2018). The ebb and flow of communication research: Seven decades of publication trends and research priorities. *Journal of Communication, 68*(2), 424–440. https://doi.org/10.1093/joc/jqx015.

Wasserman, H. (2018). Power, meaning and geopolitics: Ethics as an entry point for global communication studies. *Journal of Communication, 68*(2), 441–451. https://doi.org/10.1093/joc/jqy001.

Willems, W. (2014). Provincializing hegemonic histories of media and communication studies: Toward a genealogy of epistemic resistance in Africa. *Communication Theory, 24*(4), 415–434. https://doi.org/10.1111/comt.12043.

Youk, S., & Park, H. S. (2019). Where and what do they publish? Editors' and editorial board members' affiliated institutions and the citation counts of their endogenous publications in the field of communication. *Scientometrics, 120*(3), 1237–1260. https://doi.org/10.1007/s11192-019-03169-x.

Zhang, L., Glänzel, W., & Liang, L. M. (2009). Tracing the role of individual journals in a cross-citation network based on different indicators. *Scientometrics, 81*(3), 821–838.

Zhang, L., Janssens, F., Liang, L. M., & Glänzel, W. (2010). Journal crosscitation analysis for validation and improvement of journal-based subject classification in bibliometric research. *Scientometrics, 82*(3), 687–706.

Zhang, L., Rousseau, R., & Glänzel, W. (2016). Diversity of references as an indicator for interdisciplinarity of journals: Taking similarity between subject fields into account. *Journal of the Association for Information Science and Technology, 67*(5), 1257–1265.

Zhu, Y., & Fu, K. W. (2019). The Relationship between interdisciplinarity and journal impact factor in the field of communication during 1997–2016. *Journal of Communication, 69*(3), 273–297. https://doi.org/10.1093/joc/jqz012.

Affiliations

Manuel Goyanes[1,2] · Márton Demeter[3] · Aurea Grané[4] · Irene Albarrán-Lozano[4] · Homero Gil de Zúñiga[2,5,6]

Márton Demeter
demeter.marton@uni-nke.hu

Aurea Grané
aurea.grane@uc3m.es

Irene Albarrán-Lozano
ialbarra@est-econ.uc3m.es

Homero Gil de Zúñiga
hgz@usal.es

1 Department of Communication, Carlos III University, C/Madrid 133, Madrid, Spain

2 Democracy Research Unit (DRU), Political Science, University of Salamanca, Salamanca, Spain

3 Department of Social Communication, National University of Public Service, Budapest, Hungary

4 Department of Statistics, Universidad Carlos III de Madrid, Madrid, Spain

5 Department of Film Production and Media Studies, Pennsylvania State University, State College, USA

6 Facultad de Comunicación y Letras, Universidad Diego Portales, Santiago, Chile