



# Two-dimensional mapping of university profiles in research

Joel Emanuel Fuchs<sup>1</sup> · Thomas Heinze<sup>2</sup>

Received: 20 September 2021 / Accepted: 14 March 2022 / Published online: 9 April 2022  
© The Author(s) 2022

## Abstract

There are size-leveling indexes used to demonstrate profiling of entities in different fields, such as the Activity Index (AI) or the Index of Relative Specialization (RESP). Concentrating on the RESP, we consider German state universities as entities and their academic disciplines as fields. While it is common to illustrate several RESP values concurrently using heatmaps, we show that an interpretation of such heatmaps without further information can be misleading. Therefore, we introduce a weight for each RESP value that represents the fraction of a field at a university. Those weights correspond to RESP values that are uniquely identifiable by field and university, resulting in tuples of RESP values and their weights. We introduce a new kind of heatmap that not only illustrates RESP values but represents their corresponding weights. Those new heatmaps are less misleading than classical ones. Our introduction of a new class of heatmaps improves heatmap representation, especially for longitudinal RESP data without the need for additional tables to show the extra information.

**Keywords** Activity Index · Relative Specialization Index · German universities · German Higher Education System · Mapping of publication data · Fractional mapping · Heatmaps · Alternative heatmaps

## Introduction

Size-leveling indexes and their illustrations are very common, not only in scientometrics and bibliometrics but also in other sciences that measure fields of different-sized entities. Those indexes can be divided into two classes: based on a single variable or on multiple variables. While the field of multi-variable indexes seems to be unlimited—depending on the available variables or indicators—the field of single-variable indexes is less complex. Common examples of multi-variable indexes are values per capita, and

---

✉ Joel Emanuel Fuchs  
jfuchs@uni-wuppertal.de

Thomas Heinze  
theinze@uni-wuppertal.de

<sup>1</sup> Institute of Sociology, University of Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany

<sup>2</sup> Institute of Sociology & Interdisciplinary Center for Science and Technology Studies (IZWT), University of Wuppertal, Gaußstr. 20, 42119 Wuppertal, Germany

more complex examples are data envelopment analysis (DEA) or regression analysis (Todeschini & Baccini, 2016). One of the most important single-variable indexes is the Activity Index (AI).

Typically, the AI is applied to data sets with one numerical variable and two categorical variables (e.g., country and field for patents, or institution and subject for higher education), where each observation can be uniquely identified by a combination of two values of the categorical variables. Details for the construction of the AI are presented later. Its strength arises from its properties of (1) leveling the size of the analysis-participating entities and (2) relating the proportion of one focal field of an entity to the average proportion of that field over all entities. This enables the reader to see the importance of the focal field in the examined entity in comparison to its overall importance. Therefore, the AI is a widely used index that has generated many derivatives and thus is well probed (for example, see Aksnes et al., 2014; Fazeli-Varzaneh et al., 2020; Heinze and Fuchs, 2022; Piro et al., 2017; Teixeira et al., 2012).

Even though the AI is a well-probed instrument, there are difficulties in interpreting it. First, the AI should not be analyzed separately from its descriptive variable, because it cannot make any statement regarding the quantity of the examined variable (e.g., publications). It only describes how the examined variable is represented in a relation regarding the categorical identifiers (above or below the average of all examined variables). In the case of very low quantities in all fields of an entity, single AI values can be very high. This leads to our second point—because AI makes a statement about a relative representation (a field's share of an entity to the average share), this statement is not very clear in its interpretation. A high AI value does not imply a high value in the variable, nor a high share in the sum of this value at the focal institution, nor a high share at the focal field; one of these facts is mandatory but all can be valid. This relates to our first point. If the AI is presented attached to its original variable or something comparable, the interpretation will become easier, as we will discuss below. Therefore, it is a major deficit that heatmaps only represent the value of the AI index itself, without including more information about the basic variable. We seek to address this deficit by introducing a new type of heatmap that represents at least two levels of information on the AI, thus improving the interpretability of the heatmap.

To demonstrate an improvement in classical heatmaps (those are heatmaps representing just one numerical variable identified by two categorical variables), we introduce our method using a practical example that presents longitudinal data and its AI. We introduce publication data from the 17 technical state universities of Germany classified by the 12 main representative subject fields over the years 1995 to 2018. For this data set, we compute not only the AI but also the share of each field at each institution, needed to support the correct interpretation of AI values. The innovation of our technique is that it combines both information (AI and share of each field) in one graph. We further explain why classical heatmaps can be misinterpreted and how our two-dimensional heatmaps are more precise, which together support the need to construct such improved heatmaps.

In this process, the question arises whether another class of graphical illustration than heatmaps would be as or even more appropriate for representing the two dimensions: AI and share of a field. We address this question below, but for now, it would be presumptuous to negate this question, as it depends on the goal of the illustration. Stack area charts or stream graphs may represent the development of high and low AI parts better than our improved heatmaps, but they may reduce the clarity regarding the fields. Filled radar charts are very appropriate for representing two dimensions but are very limited in representing longitudinal data. Besides this discussion, the aim of this paper is not only to find the best

suitable graphical illustration of longitudinal AI data and the associated shares, but also to improve the illustration of heatmaps that are commonly used for AI data.

Next, we introduce our example to illustrate not only the use of the AI but also to represent a common problem in interpretation, especially when illustrating AI values by heatmaps. After introducing our example, we explain the problem caused by the interpretation of AI heatmaps, and we will address it by expanding the classical heatmap using a second dimension (i.e., information derived from the data set) to enable easier interpretation of AI heatmaps.

## AI-based interpretation of data in higher education

The webpage <https://fachprofile.uni-wuppertal.de/> presents data related to human resources, finances, bibliometric indicators, and student enrollment in the German higher education system from 1992 to 2018 (Heinze et al., 2019). This data is visualized using both the Relative Specialization Index (RSI) and the RESP, with the aim of both capturing important information and anonymizing the original data.

Mapping disciplinary profiles in higher education, especially with regard to research and teaching, has proliferated in recent years. Such mapping is done in the context of increased attention to large-scale visualization of science and technology (Borner et al., 2019; Fortunato et al., 2018). An example is Huisman et al. (2015), who classify 24 national higher education systems in Europe with regard to their degree of horizontal differentiation in research, teaching, technology transfer, and internationalization. Another example is Harzing and Giroud (2014), who identify the top-three and bottom-three research areas in 34 national higher education systems using the Revealed Comparative Advantage (RCA) measure, a derivative of the AI. Teixeira et al. (2012) used the RSI, another AI derivative, to find that private higher education entities in Portugal have research profiles complementary to those of public entities.

A well-known comprehensive mapping of bibliometric profiles was performed for Nordic universities (Piro et al., 2014, 2017). Here, the RSI is used to compare field-related publication and citation percentages with respective global field percentages, highlighting those colleges and universities with either below-average, average, or above-average contributions. In addition, Bonaccorsi et al. (2013), by means of the AI, showed for Italy that universities specializing in applied fields and engineering have a positive impact on start-ups in their region, especially in the service industries. In contrast, universities with a profile in basic science fields are related to a greater number of start-ups in manufacturing.

## The Activity Index and the Index of Relative Specialization

The AI was officially first introduced in economics by Balassa, who referred to the index as Revealed Comparative Advantage (RCA; Balassa, 1965). To honor him, sometimes the index is called Balassa index, too (Balassa, 1965). After its introduction, the AI experienced the introduction in new research fields, was renamed for this purpose several times and altered, either to be scaled for illustration issues or to reflect special cases more adequate. Frame (1977) introduced the index in bibliometrics, while Soete and Wyatt (1983) introduced the index in patent analysis naming it Revealed Technological Advantage (RTA). Schubert and Braun (1986) introduced the Attractivity Index (AAI), and therefore

were the first one who used the symmetry of the AI ( $AAI_{ij} = AI_{ji}$ ). They also were the first one who named the AI as Activity Index, even though this name was finally established by Narin et al. (1987). Naopaul and Pant (1993) introduced the term Research Priority Index (PI), to focus more on the term prioritization rather than activity. Grupp (1994) used the RTA for his patent analysis, too. Additionally, he is the first one formulating variants of the index for better illustration and analysis; namely the Revealed Patent Advantage (RPA) and the hyperbolic tangent of RPA. The last one is also known as Index of Relative Specialization (RESP; Heinze et al., 2019) or Scientific Specialization Index (SSI; Abramo et al., 2014) and is described in detail below. Even though the AI circulated as Relative Specialization Index (RSI) at least for a short time (Debackere et al., 2000), the expression RSI was coined by the European Commission (1997) meaning another scale of the AI (namely the SSI of the root square of the AI). During the period at which the expression RSI was not finalized, sometimes the AI confirmed as RSI and today's RSI as Symmetric RSI or SRSI (Murmman, 2002). We could continue the listing of introductions of the AI in different fields or the introduction of variants of the index, but we think those are enough information for a brief overview and like to continue with the definition of the AI.

Let  $D$  be a comprehensive data set with a summable variable  $v$ , where the values of  $v$  are non-negative integers. Each observation of  $D$  is uniquely identifiable by the tuple of two categorical variables  $I$  and  $J$ , e.g.,  $I$  for a set of institutions and  $J$  for a set of fields, and an observation exists for each tuple. We denote by  $v_{ij}$  the value of variable  $v$  identified by  $i \in I$  and  $j \in J$ . Then, we define the AI using Formula 1:

### Formula 1. General formula of the Activity Index (AI)

$$AI_{ij} = AI(v_{ij}) = \frac{v_{ij} / \sum_{i \in I} v_{ij}}{\sum_{j \in J} v_{ij} / \sum_{i \in I, j \in J} v_{ij}}.$$

When using longitudinal data, the AI has to be calculated for each year separately, as the formula does not consider a time variable. In most cases, one categorical variable represents entities as countries, institutions, or other groups, while the other categorical variable represents fields, subjects, or other classifications of the numerical variable  $v_{ij}$ . In the literature, it is common for the identifier  $j \in J$  to represent the entity and the identifier  $i \in I$  to represent the field. This convention is not mandatory, because the AI is symmetric (i.e.,  $A_{ij} = A_{ji}$ ).

The AI also indicates a relation of two shares. In this sense, we want to reformulate it to gain a deeper understanding of what exactly the AI is measuring. We define a share as the division of one value (e.g.,  $v_{ij}$ ) by the sum along one of the two identifiers of that value (e.g.,  $\sum_{i \in I} v_{ij}$ ; the share would be  $v_{ij} / \sum_{i \in I} v_{ij}$ ).

### Formula 2. Verbal statement of AI

$$AI_{ij} := AI(v_{ij}) = \frac{\text{The share of } v_{ij} \text{ on all accumulated observations of } j}{\text{The share of all accumulated observations of } i \text{ on all accumulated observations}}.$$

This means that the AI compares the share of field  $i$  at an institution  $j$  with the average share of field  $i$ . We can conclude some properties of the AI directly from this formula. If  $AI_{ij} > 1$ , then the share of field  $i$  at institution  $j$  is higher than the average share of field  $i$ . In contrast, if  $AI_{ij} < 1$ , then the share of field  $i$  at institution  $j$  is lower than the average share of field  $i$ .

A commonly used transformation of the AI was introduced by Grupp (1994, 1998), which we call the Index of Relative Specialization (RESP). Other names for this transformation are the Hyperbolic Tangent of the Revealed Patent Advantage (RPA<sub>h</sub>; Grupp, 1994) or the SSI (Abramo et al., 2014).

### Formula 3: Relative Specialization (RESP)

$$\text{RESP}_{ij} = 100 \frac{\text{AI}_{ij}^2 - 1}{\text{AI}_{ij}^2 + 1} = 100 \tanh(\ln(\text{AI}_{ij})) \text{ for } \text{AI}_{ij} \text{ defined as above.}$$

The RESP has two (main) advantages over the AI: it is bounded (having a range of  $[-100.0, 100.0]$ ) and symmetrical (in the sense that the AI range  $[0.0, 1.0]$  is congruent to the RESP range  $[-100.0, 0.0]$ , and the AI range  $[1.0, \infty)$  is congruent to the RESP range  $[0.0, 100.0]$ ). Therefore, it is suitable for representation in a heatmap.

## Methodology and example

At this point, we give an example of RESP values illustrated by classical heatmaps. We use the data from the project *Research and Teaching Profiles of Public Universities in Germany* found at <https://fachprofile.uni-wuppertal.de>. These data, retrieved from the Federal Statistical Office (FSO; StBA, 1992–2018), show research and teaching profiles of public universities in Germany (Heinze et al., 2019) using especially the AI or AI-based indexes to process these data. This analysis builds on a data set of 68 public universities, with information on scientific staff, basic funding, grant funding, publications, citations, and enrollment. Based on these data, institutional profiles were calculated and then visualized. Here, the visualization mainly focuses on illustrations using RSI and RESP indices. We decided to use these data as an example in this paper, not only because we are well versed in our own project, but also because the project page offers hundreds of AI-based heatmaps for download, so it was easy to retrieve a proper example. We only discuss here the structure of the subset of this data set that is used for our example instead of the complete data set.

The project obtained Web of Science (WoS) data for the 17 technical and 51 non-technical public universities in Germany. We distinguish between those two types of universities, because the structure of fields of these two groups are significantly different from each other. We decided to use the group of technical universities (namely, Brandenburgische Technische Universität Cottbus-Senftenberg, Karlsruher Institut für Technologie, Leibniz Universität Hannover, Rheinisch-Westfälische Technische Hochschule Aachen, Technische Universität Bergakademie Freiberg, Technische Universität Berlin, Technische Universität Braunschweig, Technische Universität Chemnitz, Technische Universität Clausthal, Technische Universität Darmstadt, Technische Universität Dortmund, Technische Universität Dresden, Technische Universität Hamburg, Technische Universität Ilmenau, Technische Universität Kaiserslautern, Technische Universität München, Universität Stuttgart). Further, we examined the years 1995 to 2018, since data for these years were available. The FSO defines 65 different disciplinary subjects, and every data value of the project is assigned to one certain subject. In the end, every observation is uniquely identifiable by a

**Table 1** Characteristic numbers in our data selection

No. of universities	No. of fields	No. of years	No. of publications
17	12	24	282,766

combination of institution, year, and field. For our analysis, we chose 12 research fields out of these 65; the choice is explained below.<sup>1</sup>

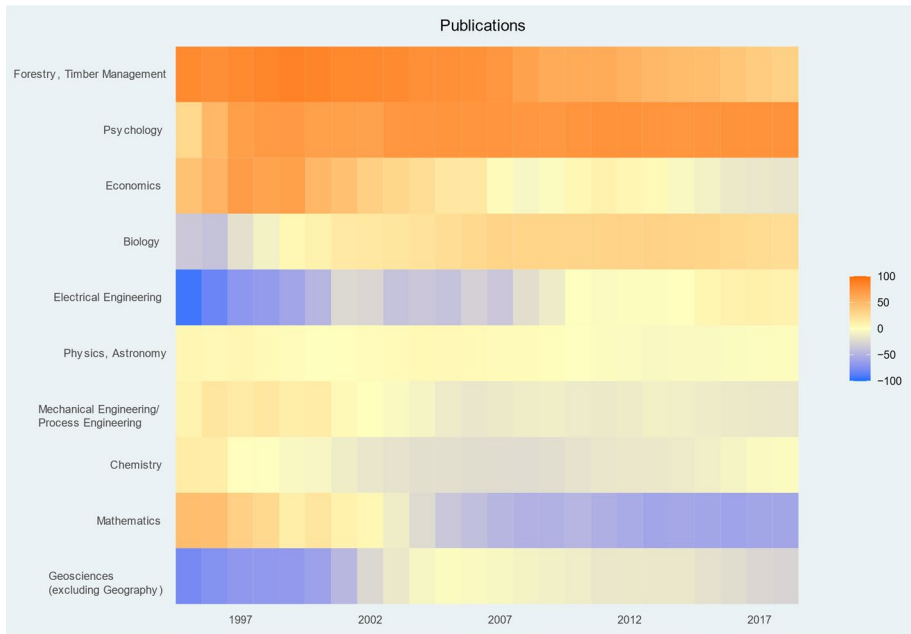
The WoS data was retrieved by the Competence Center of Bibliometrics (CCB, 2021). This data allows each entry in the WoS database to be matched clearly to one or more institutions (not only German universities, but any institution with authors/publications in WoS). Whole counting was applied, so that publication by two authors from one institution was counted as one publication from the respective institution, but publications by two authors from two institutions were counted as one publication from each of the institutions (so there are two publications). The same holds for publications with three or more authors. Besides matching each publication to one or more institutions, the CCB matched each publication to one unique field using the classification system of Archambault et al. (2011), depending on the publication journal. Heinze et al. (2019) developed a concordance table to map this classification to the classification of fields of the FSO. All this results in a clear selection and mapping of WoS-listed publications by German public universities and the fields they offer.

After attending an FSO-accordant classification of the WoS publications published by authors from German public universities, Heinze et al. (2019) examined the coverage of the FSO fields regarding the WoS database. They defined the coverage of a field by the quotient where the dividend is defined as the sum of citations made by any publication of that field and having an entry in WoS again, and the divisor is defined as the sum of citations made by any publication of that field independently of having a WoS entry.<sup>2</sup> Following this procedure, Heinze et al. (2019) identified 12 research fields at German universities that have a high coverage rate (> 50%) in WoS. These are “Biology”, “Chemistry”, “Physics, Astronomy”, “Nutrition and Household Economics”, “Psychology”, “Agricultural Sciences, Food and Beverages Technology”, “Mechanical Engineering, Process Engineering”, “Geosciences (excluding Geography)”, “Electrical Engineering”, “Forestry, Timber Management”, “Economics”, and “Mathematics”. We decided that a field with a coverage rate below 50% is not sufficiently represented by the WoS to generate any statement, so we ignored the other fields for our sample. Therefore, we considered 282,766 publications (including multiple counts resulting from authors at different institutions).

Table 1 summarizes our selection of the FSO data. These choices of FSO data are arbitrary in the sense that we could have used any real-world example to demonstrate the use of the AI and RESP index and the creation of our new type of heatmap. They are not arbitrary in the sense that we have chosen a homogeneous set of institutions and fields resulting in robust AI values. Figure 1 shows the RESP values of publications assignable to the Technical University of Dresden (TU Dresden) by a classical heatmap.

<sup>1</sup> For further details about the 51 non-technical universities and the 53 fields that were not mentioned, we point to the project’s website.

<sup>2</sup> WoS records the citations mentioned in every publication; the citation records do not automatically have their own WoS entries.



**Fig. 1** RESP values of publications at the TU Dresden (1995–2018). *Source* <https://fachprofile.uni-wuppertal.de/en.html>

### Misleading interpretation of RESP values

Figure 1 depicts the development of 10 of the 12 evaluated fields from 1995 to 2018. The missing two fields are either not offered by TU Dresden or the data of these fields are missing. The fields are sorted from top to bottom by the sum of their RESP values over all 24 years. Even if the heatmap shows longitudinal data, we want to focus on the year 2018 and additionally present some numbers.

As mentioned above, the RESP values for “Psychology”, “Forestry, Timber Management”, and “Biology” are far in excess of the average (RESP=0), while the RESP values for “Physics, Astronomy” and “Chemistry” are positioned near the average. What does this mean? As Formula 2 tells us, this means that the shares of publications in “Psychology”, “Forestry, Timber Management”, and “Biology” in all publications at TU Dresden are above the average shares of these fields at all 17 TUs. In contrast, the shares of publications in “Physics, Astronomy” and “Chemistry” in all publications from TU Dresden are around the average shares of these fields at all 17 TUs. These statements do not mean that the three fields having high RESP values are of particular importance for TU Dresden. We observed in conversations and discussions with users of our project site, and also with readers not familiar with statistics but interested in higher education research, that these users often associate high RESP values with high efforts in the corresponding fields. This is a common misinterpretation of AI and RESP values. Additionally, it is well known that rising absolute values do not mean rising AI values and vice versa (Rousseau & Yang, 2012), and scaling by derivatives like RSI and RESP may be arbitrary (Stare & Kejžar, 2014). All things considered, a second value supporting an AI interpretation is necessary. Table 3 shows the

**Table 2** RESP values from TU Dresden in 2018

Field (FOS schema)	RESP value
Psychology	76.56
Forestry, Timber Management	33.02
Biology	24.56
Electrical Engineering	9.01
Physics, Astronomy	−1.58
Chemistry	−2.34
Mechanical Engineering, Process Engineering	−15.45
Economics	−16.74
Geosciences (excluding Geography)	−27.60
Mathematics	−58.66

**Table 3** Publications of presented fields in 2018

Field	Publications at TU Dresden	Share at TU Dresden (%)	Publications at all 17 TUs	Share at all 17 TUs (%)
Psychology	84	4.2	356	1.9
Forestry, Timber Management	15	0.7	85	0.4
Biology	249	12.4	1930	10.2
Physics, Astronomy	652	32.4	5427	28.6
Chemistry	386	19.2	3320	17.5
All Publications	2014	100.0	18,966	100.0

absolute numbers for the corresponding RESP values and a more detailed picture of these fields.

Examining “Forestry, Timber Management”, we can identify two different levels of interpretation. In Table 3, we call these two levels the row level of interpretation (considering only publications in “Forestry, Timber Management” from all 17 TUs) and the column level of interpretation (considering publications of all fields, but only at TU Dresden). At the row level, we have 15 publications at TU Dresden, while there are 70 (= 85–15) publications at the other 16 TUs. More than every sixth publication covering the field “Forestry, Timber Management” was (co-)written by an author from TU Dresden (precisely 17.6%). This explains the high RESP value of 33.02 (see Table 2) and indicates that authors and therefore scientists from TU Dresden are important contributors to this field.

We get a different picture if we interpret the column level. In 2018, authors from TU Dresden covered the 12 fields in our data set by 2014 publications; only 15 of them handled the field “Forestry, Timber Management”. From this point of view, it seems that the field “Forestry, Timber Management” is less productive considering the whole contingent of publications from TU Dresden. It seems to be a niche field, perhaps profiling TU Dresden but not representing it or its publication behavior. Just considering the quantity of publications, the field “Physics, Astronomy” is most important to TU Dresden, as every third publication of TU Dresden is in this field. Even though “Physics, Astronomy” plays such a major role at TU Dresden, it is not a highlighting characteristic, because the average share of publications handling this field is almost as high as the share at TU Dresden (28.6% on average to 32.4% at TU Dresden; see Table 3).



Hence, each level on its own induces occasions for misinterpreting the data. If we want to analyze both levels properly, we must examine the two levels together. The RESP value of “Forestry, Timber Management” is 33.02, and its share regarding all publications from TU Dresden is 0.7%; in brief, the RESP tuple is (33.02; 0.7%). Now, the reader immediately has the information that TU Dresden plays a major role in the field of “Forestry, Timber Management” (because of the RESP value of 33.02), but that the field “Forestry, Timber Management” plays no major role at TU Dresden (because its share is only 0.7%). Analogous to this, the RESP tuple (− 1.58; 32.4%) for “Physics, Astronomy” at TU Dresden tells us that “Physics, Astronomy” plays a major role at that institution (based on its share of 32.4%), but TU Dresden plays only an average role in the field of “Physics, Astronomy” (based on the RESP value of − 1.58).

### Weighting RESP values and a new representation of heatmaps

With the introduction of RESP tuples, the issue of illustration arises. RESP values are often represented not in tables but in heatmaps; for example, see Heinze et al. (2019) or Piro et al. (2017). Heatmaps were originally created to contain three different levels of information: two categorical variables on the *x*-axis and *y*-axis that uniquely identify an observation, and a color that represents the value of the observation. As in Fig. 1, the *x*- and *y*-axes allow the RESP value of a certain field in a certain year for TU Dresden to be identified, while the color in the corresponding cell represents the RESP value. However, classical heatmaps are not intended to include a fourth level of information needed to represent a tuple consisting of two dimensions, such as RESP value and share.

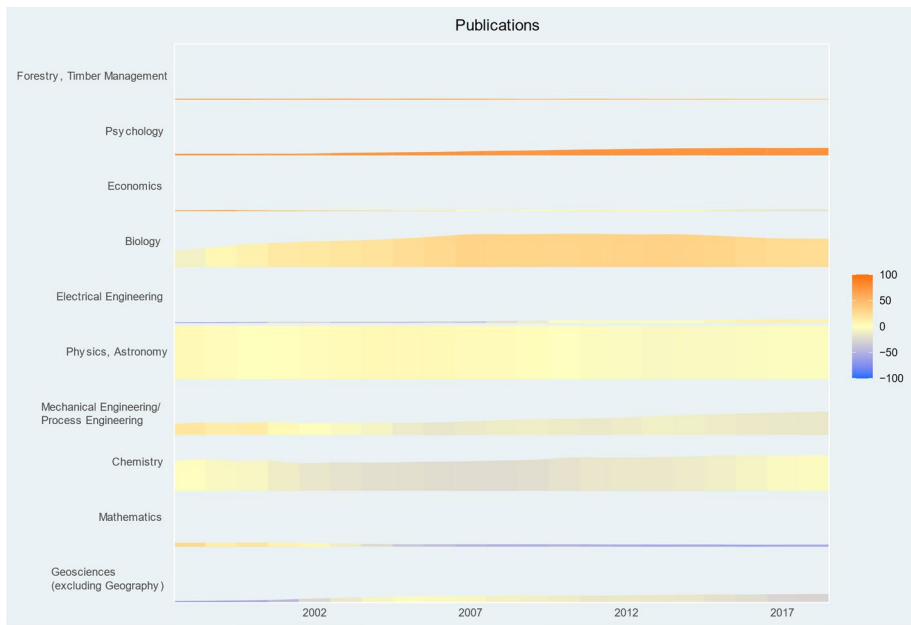
There are two obvious solutions to address this problem. First, we could supply every heatmap with a corresponding table containing the missing tuples. That would cost (valuable) space in a paper or publication and would contraindicate use of a heatmap in the first place. Second, we could integrate the missing information (the share of a field at an institution) into the heatmap. Suppose each box of a heatmap is one unit high. It is possible to vary this height so that it corresponds to the share of the field represented by the affected box. That is, in the heatmap of all RESP values, the cell for “Physics, Astronomy” would have a height not of one unit but of 0.324 units (representing 32.4%), and, in the same way, the cell for “Forestry, Timber Management” would have a height of 0.007 units (representing 0.7%).

This would be quite an easy way to integrate RESP tuples into heatmaps, but it is also not feasible. Observing 12 fields, the expected average share is around 8%, so most boxes would have a height of about 0.08 units. Heatmaps consisting of cells or bars with such an average height would show only an empty grid. Therefore, instead of scaling the height of a box to the corresponding share of the RESP tuple, we first weight all tuples, more precisely the share value of the RESP tuples.

Let  $(RESP_{ij}, S_{ij})$  be an RESP tuple of field *i* and institution *j*.  $S_{ij}$  is the share of the number of publications in field *i* relative to the number of all publications from the institution *j*, as given by the formula  $S_{ij} = \frac{v_{ij}}{\sum_{i \in I} v_{ij}}$ . Then, the weighted RESP tuple  $(RESP_{ij}, wS_{ij})$  is defined as follows:

**Table 4** Weighted RESP tuples for TU Dresden

Field	Publications	Shares (%)	RESP values	Weighted RESP tuples
Psychology	84	4.2	76.56	(76.56; 0.13)
Forestry, Timber Management	15	0.7	33.02	(33.02; 0.02)
Biology	249	12.4	24.56	(24.56; 0.38)
Physics, Astronomy	652	32.4	−1.58	(−1.58; 1.00)
Chemistry	386	19.2	−2.34	(−2.34; 0.59)
All Publications	2014	100.0		

**Fig. 2** Two-dimensional mapping of weighted RESP tuples of publications at TU Dresden (1998–2018)**Formula 4. Weighted RESP tuple**

$$\left( \text{RESP}_{ij}, wS_{ij} \right) = \left( \text{RESP}_{ij}, \frac{v_{ij}}{\max_{i \in I} v_{ij}} \right) = \left( \text{RESP}_{ij}, \frac{S_{ij}}{\max_{i \in I} S_{ij}} \right).$$

Formula 4 demonstrates that we are weighting the shares of all fields at one institution by the maximum share at this institution. In our example, we achieve this for TU Dresden by dividing all shares by the maximum share of 32.4%, which is the share of publications in the field “Physics, Astronomy” relative to all publications from TU Dresden. Table 4 repeats the results of Table 3, but instead of the columns showing the aggregation of all 17 TUs, it shows the weighted RESP tuples.

As Table 4 shows, weighted shares are shown in the full range 0 to 1 unit; thus, they can be used as the height of the corresponding heatmap cells. We computed the weighted

RESP tuples for all 10 fields and 24 years of the publications from TU Dresden and illustrated the RESP tuples as a first draft. We noticed that for the new representation of RESP tuples, the staccato-like course from one year to the next within a field is difficult to view, so we smoothed all heatmap cells for one field over the years. For this purpose, all annual values are placed in the middle of the bar, and adjacent bars are connected using a natural cubic spline interpolation, which has the effect of smoothing the bar graph. The result is presented in Fig. 2. We call the new heatmap in Fig. 2 “two-dimensional”, because it represents not only RESP values but the weighted RESP tuples. Here, the first “dimension” is the colored representation of the RESP values, while the second “dimension” is the height-sized representation of the share of fields (weighted by the maximum share of the corresponding year and the focal institution).

## Results and discussion

As with Figs. 1 and 2 shows that publications from TU Dresden in the fields “Forestry, Timber Management” and “Psychology” play a major role in these fields. This can be visually observed from the deep-orange color of the two fields (equivalent to a RESP value near 100). Simultaneously, we see that these two fields do not play a major role at TU Dresden, because the heights of their bars are pretty low. This is clearly new information, as Fig. 1 implied that these two fields are leading fields at TU Dresden. The most important field at TU Dresden (measured by having the largest share of publications from TU Dresden) is the field “Physics, Astronomy”, as shown in Fig. 2, because the bar for “Physics, Astronomy” has a constant height of 1 unit. Even if “Physics, Astronomy” is the field with the largest share, its RESP values are only represented by the color yellow, meaning they range around 0. Consequently, the share of publications in the field “Physics, Astronomy” at TU Dresden is seen to be only average.

If we want to identify a profiling field at TU Dresden from Fig. 2, it would most likely be “Biology”. Its heatmap color is light orange (equivalent to a RESP value around 50), indicating that publications at TU Dresden play a more important role in this field than at other universities (always measured in the sense of quantity of publications in this field in comparison to all publications at a university). Simultaneously, its height indicates that the third largest share of publications at TU Dresden is located in this field. This is also reflected by the RESP tuples (see also Table 4), but Fig. 2 helps us easily observe the 240 illustrated RESP tuples in comparison to using a table.

Summing up, we introduced a two-dimensional graphical mapping with an emphasis on research profiles of public universities. The new graphical representation can also be applied to other topics, such as teaching, technology transfer, or internationalization. The key difference compared to existing heatmaps is that our new graphs capture comparisons for both the nationwide level (“between” in our example represented by the question of how TU Dresden “performs” in a field) and the organizational level (“within” in our example represented by the question of how a field “performs” at TU Dresden). In this way, we take a first step toward better understanding of the interplay between both levels.

Regarding the graphical representation in Fig. 2, one might wonder about other technical possibilities, such as Sankey diagrams, stack area charts, or stream graphs (Wickham, 2016). Even though this paper focuses on heatmaps, we also want to examine those other solutions briefly. We only consider graphs capable of representing the same information as our new class of heatmaps. Of course, there are other possibilities for illustrating AI and

RESP values, such as scatter plots and histograms, but they illustrate other information with another aim. Our new class of heatmaps represents longitudinal data for a given institution, namely years on the  $x$ -axis, fields on the  $y$ -axis, the RESP value as the color of each  $x$ - $y$  cell, and finally the proportion of the field at the institution as the height of the  $x$ - $y$  cell. (The fact that we smoothed the cell heights to please the reader's eye does not contribute any information.)

Any competing graph must also map the two initial pieces of information for year and field to the two desired pieces of information for RESP value and field proportion, all regarding a given institution. Stack area charts can do this, but after stacking the fields, it is very difficult for the reader to locate a specific year–field combination in the illustration. Sankey diagrams are totally improper for this task, because they show the flow of resources from one category to another. The AI values and also the field proportions are not capable of providing the information regarding which other fields received decreasing resources. Stream graphs are very similar to the stacked case. Of course, if we color each stream year-by-year corresponding to its RESP value, it could illustrate all required information, but again, the result would be too confusing to easily show certain year–field values.

Because the aim of this paper is to improve widely used heatmaps, we did not examine all other graphical solutions. In view of these three very popular solutions, we close the discussion on alternative representations to our new class of heatmaps, although we still have one promising candidate: the spider diagram. For a given institution, it can represent all fields and the proportion of each field. In addition, the area of the spider diagram can be filled to correspond to the RESP values of each field. The result is very pleasing to the reader's eye, but unfortunately one piece of information is lost: the year. Therefore, this solution is only useful for a given institution and a given year.

Two words of caution are in order. First, we are aware that our contribution is descriptive and that further statistical analyses regarding the two levels and their roles in both building and maintaining institutional profiles are necessary. Second, we examined relations based on quantities of publications at institutions. The statements we are gathering are the relative assembly of fields at a university in a nationwide comparison. Our examinations do not state anything about the “real” performance of universities.

Unfortunately, the scope of applications for our new heatmap class is very narrow, as it only refers to the illustration of longitudinal AI and RESP values. We believe this is not due to a lack of need for an analysis of longitudinal AI data, but to a lack of illustration possibilities. Generally, and especially in higher education research, a global development of key figures such as finance and staff can logically be enriched by the AI development for each institution. We expect that such an analysis will be performed in the future. Regarding our own project, we will implement the new heatmap class as standard and observe how well users and researchers accept the new illustration.

## Notes

The analysis was conducted in R (R Core Team, 2020) with `data.table` (Dowle & Srinivasan, 2019), and the figures were produced using `ggplot2` (Wickham, 2016). For smoothing the heat bars, the packages `stats` and `splines` (R Core Team, 2020) were used. The pseudo code for smoothing is (as a geometric part of the `ggplot` command):

```
stat_smooth(method = 'glm', method.args = list(family = gaussian),  
formula = y ~ splines::ns(x, df = years - 1), se = FALSE, geom = "ribbon", span = 1).
```

**Acknowledgements** This paper was originally part of the proceedings of the 18th International Conference on Scientometrics and Informetrics (ISSI 2021) as a full paper (Fuchs & Heinze, 2021). It is a substantially reworked version of the one presented at the ISSI 2021, and at least 25% of the pages contain new material. We thank Dirk Tunger and Paul Eberhardt for their support in data collection and curation.

**Funding** Open Access funding enabled and organized by Projekt DEAL.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## References

- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2014). A new bibliometric approach to assess the scientific specialization of regions. *Research Evaluation*, 23(2), 183–194.
- Aksnes, D. W., van Leeuwen, T. N., & Sivertsen, G. (2014). The effect of booming countries on changes in the relative specialization index (RSI) on country level. *Scientometrics*, 101, 1391–1401. <https://doi.org/10.1007/s11192-014-1245-3>
- Archambault, E., Beauchesne, O., & Caruso, J. (2011). Towards a multilingual, comprehensive and open scientific journal ontology. In *Proceedings of the 13th international conference of the International Society for Scientometrics and Informetrics* (pp. 66–77).
- Balassa, B. (1965). Trade liberalization and “revealed” comparative advantage. *The Manchester School of Economic and Social Studies*, 32, 99–123.
- Bonaccorsi, A., Colombo, M. G., Guerini, M., & Rossi-Lamastra, C. (2013). University specialization and new firm creation across industries. *Small Business Economics*, 41, 837–863.
- Borner, K., Bueckle, A., & Ginda, M. (2019). Data visualization literacy: Definitions, conceptual frameworks, exercises, and assessments. *Proceedings of the National Academy of Sciences of the United States of America*, 116(6), 1857–1864. <https://doi.org/10.1073/pnas.1807180116>
- CCB. (2021, November 13). Competence Center of Bibliometrics. <https://www.bibliometrie.info/en/>
- Debackere, K., Luwel, M., & Veugelers, R. (2000). *Patent data as a tool to monitor S and T portfolio's* (pp. 1–20). K.U. Leuven-Departement toegepaste economische wetenschappen.
- Dowle, M., & Srinivasan, A. (2019). data.table: Extension of data.frame. R package version 1.12.8. <https://cran.r-project.org/package=data.table>
- European Commission. (1997). *EUR 17639—Second European Report on S&T Indicators 1997*. Office of the Official Publications of the European Communities.
- Fazeli-Varzaneh, M., Noruzi, A., Noroozi Chakoli, A., & Sarrafzadeh, M. (2020). The national and international comparison of relative specialization, citation and cooperation indicators of Iran in water resources research. *Scientometrics Research Journal*. <https://doi.org/10.22070/rsci.2020.5673.1416>
- Fortunato, S., Bergstrom, C. T., Borner, K., Evans, J. A., Helbing, D., Milojevic, S., & Barabasi, A. L. (2018). Science of science. *Science*. <https://doi.org/10.1126/science.aa0185>
- Frame, J. D. (1977). Mainstream research in Latin America and the Caribbean. *Interciencia*, 2, 143–148.
- Fuchs, J. E., & Heinze, T. (2021). Two-dimensional mapping of university profiles in research. In *18th International conference on scientometrics and informetrics ISSI2021. Proceedings* (pp. 425–434).
- Grupp, H. (1994). The measurement of technical performance of innovations by technometrics and its impact on established technology indicators. *Research Policy*, 23, 175–193.
- Grupp, H. (1998). Measurement with patent and bibliometric indicators. In H. Grupp (Ed.), *Foundations of the economics of innovation. Theory, measurement, practice* (pp. 141–188). Edward Elgar.
- Harzing, A.-W., & Giroud, A. (2014). The competitive advantage of nations. An application to academia. *Journal of Informetrics*, 8, 29–42.
- Heinze, T., & Fuchs, J. E. (2022). National and organizational patterns of Nobel Laureate careers in physiology/medicine, physics, and chemistry. *Scientometrics*. <https://doi.org/10.1007/s11192-021-04250-0>

- Heinze, T., Tunger, D., Fuchs, J. E., Jappe, A., & Eberhardt, P. (2019). *Research and teaching profiles of public universities in Germany. A mapping of selected fields*. BUW. <https://doi.org/10.25926/9242-ws58>.
- Heinze, T., Tunger, D., Fuchs, J. E., Jappe, A., & Eberhardt, P. (2021). *Research and teaching profiles of public universities in Germany*. Chair of Organizational Sociology. Wuppertal: BUW. <https://fachp-rofile.uni-wuppertal.de/>
- Huisman, J., Lepori, B., Seeber, M., Frölich, N., & Scordato, L. (2015). Measuring institutional diversity across higher education systems. *Research Evaluation*, 24, 369–279.
- Murmann, J. P. (2002). The coevolution of industries and national institutions: Theory and evidence. Discussion Papers/Wissenschaftszentrum Berlin für Sozialforschung, Forschungsschwerpunkt Markt und politische Ökonomie 02-14. Wissenschaftszentrum Berlin für Sozialforschung gGmbH.
- Naopaul, P. S., & Pant, N. (1993). Research priorities of major countries in artificial intelligence. In A. Ghosal & P. N. Murthy (Eds.), *Recent advances in cybernetics and systems*. Tata McGraw-Hill.
- Narin, F., Carpenter, M. P., & Woolf, P. (1987). Technological assessments based on patents and patent citations. In H. Grupp (Ed.), *Problems of measuring technological change* (pp. 107–119). TÜV Rheinland.
- Piro, F. N., Aldberg, H., Aksnes, D. W., Staffan, K., Leino, Y., Nuutinen, A., & Sivertsen, G. (2017). *Comparing research at Nordic higher education institutions using bibliometric indicators covering the years 1999–2014. Policy Paper 4/2017*. NIFU.
- Piro, F. N., Aldberg, H., Finnbjörnsson, P., Gunnarsdóttir, O., Karlsson, S., Skytte Larsen, K., & Sivertsen, G. (2014). *Comparing Research at Nordic Universities using Bibliometric Indicators—Second report, covering the years 2000–2012. Policy Paper 2/2014*. NordForsk.
- R Core Team. (2020). *R: A language and environment for statistical computing*. Retrieved from Vienna <https://www.r-project.org/>
- Rousseau, R., & Yang, L. (2012). Reflections on the activity index and related indicators. *Journal of Informetrics*, 6(3), 413–421. <https://doi.org/10.1016/j.joi.2012.01.004>
- Schubert, A., & Braun, T. (1986). Relative indicators and relational charts for comparative assessment of publication output and citation impact. *Scientometrics*, 9, 281–291.
- Soete, L. G., & Wyatt, S. M. E. (1983). The use of foreign patenting as an internationally comparable science and technology output indicator. *Scientometrics*, 5, 31–54.
- Stare, J., & Kejžar, N. (2014). On standardization of the Activity Index. *Journal of Informetrics*, 8(3), 503–507. <https://doi.org/10.1016/j.joi.2014.04.004>
- StBA. (1992–2018). *Bildung Und Kultur. Fachserie 11*. Statistisches Bundesamt.
- Teixeira, P. N., Rocha, V., Biscaia, R., & Cardoso, M. F. (2012). Competition and diversity in higher education: An empirical approach to specialization patterns of Portuguese institutions. *Higher Education*, 63(3), 337–352.
- Todeschini, R., & Baccini, A. (2016). *Handbook of bibliometric indicators: Quantitative tools for studying and evaluating research*. Wiley.
- Wickham, H. (2016). *ggplot2: Elegant graphics for data analysis*. Springer.