INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM

Working Paper Series

Using the Leiden Rankings as a Heuristics: Evidence from Italian universities in the European landscape

Cinzia Daraio [a]
Simone Di Leo [a]
Loet Leydesdorff [b]

[a] DIAG Sapienza University of Rome, Italy.
[b] University of Amsterdam, Amsterdam School of Communication Research, The Netherlands.

2022/08 **March 2022**
ISSN(ONLINE) 2284-0400

Using the Leiden Rankings as a Heuristics: Evidence from Italian universities in the European landscape[a]

Cinzia Daraio, Simone Di Leo,* and Loet Leydesdorff***

*DIAG Sapienza University of Rome (Italy), Email: daraio@diag.uniroma1.it, dileo@diag.uniroma1.it

**University of Amsterdam, Amsterdam School of Communication Research (The Netherlands), Email: loet@leydesdorff.net

Abstract

We propose an innovative use of the information provided by the Leiden Rankings (LR). Although LR only consider research output of major universities reported in Web of Science, statistical analysis of LRs combined with network mapping can reveal the complexity of research performance measurement. Yet, one can identify "outlying" institutions that perform significantly below or above expectations; these can be further analyzed using case studies. Outliers can inform and guide science policies about alternative options. Analyzing the case of the Politecnico di Bari, we observe that "small teams" led by young and promising scholars can push the performance of a university up to the top of the LR. Supporting "emerging teams", as argued by Moed (2017), can thus be an alternative to research support policies, adopted to encourage virtuous behaviors and best practices in research.

Keywords: university rankings, Leiden Rankings, Italian universities, best research practices, Politecnico di Bari, case study, emerging groups

[a] Some of the ideas of this work were first discussed with Henk F. Moed. After his sudden passing, we decided to complete this paper and dedicate it to his memory.

1. Introduction

Universities are subject to constant scrutiny by various stakeholders. The "evaluation society" (Dahler-Larsen, 2011) in which we live, did not spare universities. Universities must be transparent and accountable for the public money invested in their activities. However, the evaluation of performance by universities is far from simple. The multiple activities that universities carry out, which include teaching, research, and the "third mission", interact with one another and with the objectives of policy makers and institutional missions.

Rankings have been established as a tool for informing the governance of universities. Rankings, however, influence institutional behavior and increase the competition in the higher education system, asking for policy measures in response to rankings. Even today, some twenty years after the introduction of the first rankings, there is still a great deal of interest in the theory and methodology of rankings and their impact and influence. (see e. g. Hazelkorn and Mihut, 2021).

In a number of studies the inconsistencies of the rankings have been analyzed from a methodological perspective. For example, Fauzi et al. (2020) analyzed five of the leading world rankings which include: Quacquarelli Symonds (QS), Times Higher Education (THE), Academic Ranking of World Universities (ARWU), Leiden Rankings (LR), and Webometrics ranking. Similarly, Olcay and Bulu (2017) analyzed the following rankings: University Ranking by Academic Performance (URAP), THE, ARWU, QS and LR. Moed (2017) compared the ARWU, LR, THE, QS and U-Multirank rankings. These authors also analyzed the geographic coverage, the overlap of the institutions across rankings, and how the indicators were calculated from the raw data.

Moed (2017) argued that existing rankings provide purposeful information on a single aspect rather than a multidimensional information system. In response to this critique, Daraio et al. (2015) proposed a new approach that allows to overcome the four main limitations of university rankings, namely: (1) mono-dimensionality; (2) lack of robustness from a statistical point of view; (3) dependence on the scale and subject specializations of universities; and (4) the absence of considerations of the input–output structure of academic activities. These authors proposed to rank universities based on the integration of different kinds of information and the use of more robust ranking techniques, based on advanced nonparametric efficiency methods. Moed and Halevi (2015) proposed a multidimensional matrix of scientific indicators to support the choice of metrics to be applied in a research evaluation process depending on the unit of evaluation, the dimension of the research to be evaluated, and the aims and policy contexts of a specific evaluation.

On the basis of an analysis of the main problems of existing rankings, Daraio and Bonaccorsi (2017), furthermore, proposed to invest in a data infrastructure instead of investing heavily in the creation of integrated datasets for specific indicators. Data may soon prove obsolete given the need for policy makers to have more and more granular and contextualized indicators available. Given the dynamic of the landscape, these authors proposed to co-produce indicators within open data platforms "combining heterogeneous sources of data to generate indicators that address a variety of user requirements" (Daraio and Bonaccorsi, 2017, p. 508). Vernon et al. (2018) claimed that no single ranking system provides a comprehensive assessment of the quality of academic research and stated that the measurement of university research performance through standardized rankings remains a candidate to be further investigated.

In the next sections, we elaborate on the information provided by some rankings, including LR. Is it possible to provide higher education institutions with useful feedback on their scientific achievements in order to improve their respective performances? Following Leydesdorff *et al.* (2019), we propose the use of LR to cluster universities into groups of universities which are not significantly different in terms of relevant statistics. Universities that are not statistically different in terms of their output can be considered as belonging to the same group. For example, these authors used the "excellence indicator" PPtop-10%—that is, the proportion of the top-10% most-highly-cited papers assigned to a specific unit of analysis (e.g, university; cf. Waltman, et al., 2012). This is a percentile indicator and therefore size-independent. Percentiles can be used as an alternative of normalized citation impact of scientific publications. Bornmann et al. (2013) reviewed advantages and limitations of percentile ranks in bibliometrics. Leydesdorff *et al.* (2019) thus analyzed 902 universities in 54 countries; focusing on UK, Germany, Brazil, and the USA. Applying the same methodology, Leydesdorff *et al.* (2021) analyzed and compared 205 Chinese universities with 197 US universities in LR 2020.

Behind the production of the LR there is a huge standardization and data cleaning work on the names of the organizations. This activity, pioneered by Henk F. Moed, led to the development of an extensive and sophisticated system to identify, register, and harmonize organization names (Calero-Medina et al. 2020).In this study, we consider options to use LR as a heuristics to investigate outlier institutions. This means to elaborate on LR to identify outlying universities, i.e. those that outperform expected results. Once the outliers have been identified, through statistical analysis and mappings, one can carry out case studies on individual successful institutions. By using the bibliometric information derived from the LR as a database, it is possible to further characterize the topics specialization and the research organization of the investigated outliers. We show how this new use of the LR can work in practice by elaborating the analysis for the Politecnico di Bari as a case.

The paper is organized as follows. In the next section we describe the main objective of the work and its contribution to the literature. Section 3 describes the methods used for the analysis carried out, while Section 4 illustrates data used in the empirical analysis. Section 5 reports the main results, Section 6 discusses the obtained results, and Section 7 concludes the paper. Supplementary materials are reported in appendix.

2. Aims and contribution

The main objective of this work is to propose an innovative use of the information provided by the Leiden Rankings (LR). We propose to use the LR as a source for identifying heuristics in the performance of universities to further investigate in details through case studies. The object of the heuristics are outlier institutions defined as institutions that outperform expected results, i.e. institutions performing significantly different from the expectation. Although LR only provide information on the research output of major research universities, analyzing LR through statistical analysis (Leydesdorff et al., 2019) and network mapping (Blondel et al. 2008; Van Eck and Waltman, 2010), can be useful for delving into the complex performance of universities.

We focus on the position of Italian universities in the European scientific landscape, on the basis of their respective values as depicted by the LR. Italy provides an interesting international case study as the Italian academic system is primarily a public system, but has one of the lowest public funding rates in Europe. Nonetheless, it has levels of scientific production (measured in terms of published articles) and scientific impact (measured by the number of citations received) comparable to the majority of other countries with a similar level of economic development. In Italy, two thirds of the funds are allocated to universities on the basis of the numbers of students enrolled at the university and the remainder on the basis of scientific production weighted with the quality of research.

Despite the exclusive focus of LR on the output of the research insofar as indexed in Web of Science, we show how LR can be used to make a representation of the positioning of the different universities from which to identify outlier institutions as heuristics on which carry out in-depth case studies. This will be showed in the case of the Politecnico di Bari that is the only polytechnic in the South of Italy and also the youngest polytechnic in Italy. Politecnico di Bari ranks first not only in Italy but also in Europe in the Social Science and Humanities (named SSH in the following) research field, and it is second worldwide.

We aim to contribute to the current research activity on rankings, their limits and their potential, linking this research with the research line on teams (Guimera et al. 2005, Wuchty et al. 2007, and Wu et al. 2019) and on good research practices (Daraio and Vaccari, 2020). In our research "teams" are identified as an expectation on the basis of previous research but also a result of the analysis. Finally, we discuss the implications of the results in terms of research policies related to alternative ways to allocate research funds (Ioannidis, 2011; Stephan, 2012; Moed, 2017).

3. Data

Data of LR 2021 were downloaded in Excel format from http://www.leidenranking.com/downloads. LR 2021 analyzed 902 universities in 69 countries.

The file contains ranks for these universities in the preceding years (in intervals of four years). Rankings are counted both fractionally and in whole numbers. Data is provided for "All sciences" and five major fields: (*i*) biomedical and health sciences (BIO), (*ii*) life and earth sciences (LIFE), (*iii*) mathematics and computer science (MAT), (*iv*) physical sciences and engineering (FIS), (*v*) social sciences and humanities (SSH).

First, we explored "All sciences" (cf. Strotmann and Zhao, 2015), the last available period (2016-2019), and fractional counting. Only the fully-covered core journals and not the non-core journals are included. Thereafter we analyze also the five major fields of science as distinguished in LR. If so wished, the analysis can be repeated analogously with differently classified data. See SM 1 for additional information and for an available routine to extract the data.

4. Methods

The methods used in this paper are based on the approach used by Leydesdorff *et al.* (2019) combined with network and mapping techniques (Blondel et al. 2008; Van Eck and Waltman, 2010). Leydesdorff *et al.* (2019) compared three methodologies for the identification of homogeneous groups vs. statistically significant differences using the PP-top10% as the dependent variable: (1) *stability intervals* (e.g., Colliander and Ahlgren 2011, at p. 105), (2) the *z*-test which is based on the chi-square distribution (Leydesdorff and Bornmann, 2012), and (iii) *power analysis* (Cohen, 1977). The conclusion of the comparison between UK and German universities was that the first two methods provided comparable results, but the third one (that is, power analysis) led to very different results, which the authors were not able to explain. In order to focus on the substantial research question about Italian universities, we limit the discussion to using the *z*-test for the groupings.

We are aware of the limits of the use of statistical significance tests in research assessment. However, given the exploratory nature of our approach which looks to identify outliers institutions to further analyze by means of in-depth case studies, we are less affected by the limits identified in literature (see e.g. Schneider, 2013). We use the Louvain-algorithm for community finding (Blondel *et al*., 2008), because it provides less isolates than the algorithm of VOSViewer. Note that we use VOSviewer for the visualization (cf. Abramo, d'Angelo, and Grilli, 2016), but not for the decomposition. We complement these methods with descriptive and factorial analyses.

A numerical example

The *z*-test can be used to measure the extent to which an observed proportion differs significantly from expectation. In the case of $PP_{top\ 10\%}$, the expectation is 10%: without prior knowledge, one can expect a randomly selected sample to contain 10% of publications in the top 10%. The test statistics can be formulated as follows:

$$z = \frac{p_1 - p_2}{\sqrt{p(1-p)\left[\frac{1}{n_1} + \frac{1}{n_2}\right]}} \tag{1}$$

where: n_1 and n_2 are the numbers of all the papers published by institutions 1 and 2 (under the column "*P*" in LR); and p_1 and p_2 are the values of $PP_{top\ 10\%}$ of institutions 1 and 2. The pooled estimate for proportion (p) is defined as:

$$p = \frac{t_1 + t_2}{n_1 + n_2} \tag{2}$$

where: t_1 and t_2 are the numbers of top-10% papers of institutions 1 and 2. These numbers can be calculated on the basis of "*P*" and "$PP_{top\ 10\%}$.". When testing values for a single university, $n_1 = n_2$, p_1 is the value of the $PP_{top\ 10\%}$, $p_2 = 0.1$, and $t_2 = 0.1 * n_2$ (that is, the expected number in the top-10%).

An absolute value of z larger than 1.96 indicates statistical significance of the difference between two ratings at the five percent level ($p < 0.05$). The threshold value for a test at the one percent level ($p < 0.01$) is 2.576; $|z| > 3.29$ for $p < 0.001$. In a series of tests for many institutions, one may wish to avoid family-wise accumulation of Type-I errors by using the Bonferroni correction; that is, $p_{\text{Bonferroni}} = \alpha / n$ where α is the original test-statistics and *n* the number of comparisons.

Universities which are not statistically significantly different can be considered as belonging to the same performance group. Despite differences in $PP_{top\ 10\%}$ the performance of these universities can be denoted as similar in statistical terms. As noted above, this group membership is represented as links of a network, so that groups can be visualized and analyzed using network software.

At http://www.leydesdorff.net/leiden11/index.htm the user can retrieve a file leiden11.xls which allows for feeding *P* and $PP_{top\ 10\%}$ values harvested from the LR for each two universities. The spreadsheet provides the significance level of the difference measured as *z*-score. For example, Politecnico di Bari is listed (in the category "All sciences" of LR 2021) with $P = 800$ articles of which 112 (14%) participate in the top-10% layer for the comparable set worldwide ($PP_{top\ 10\%}$); the upper and lower bounds are 11.9 and 16; $z = 6.29$. The Politecnico di Milano has 4268 articles with $PP_{top\ 10\%}$ = 11.5%, bounded between 10.8 and 12.3. The stability intervals are thus intersecting. For the *z*-test one needs the pooled estimate:

$$\hat{p} = \frac{(112.00 + 490.82)}{(800+4268)} = 0.1189 \qquad (2)$$

Using Equation 1 (above), it follows that z = 2.004. The difference between Bari and Milano is thus statistically significant at the 5% level.

5. Results

5.1. ***Clusters of European Universities***

We analysed 302 European Universities included in the LR of 2021 employing a multi-level Louvain Communities detection approach (see Blondel et al., 2008). Fig. 1 shows the three clusters obtained on the European universities by applying the approach of Blondel et al. (2008). We observe a strong presence of Dutch and Belgium universities. North-western Europe (including Denmark; part of Germany) are central.

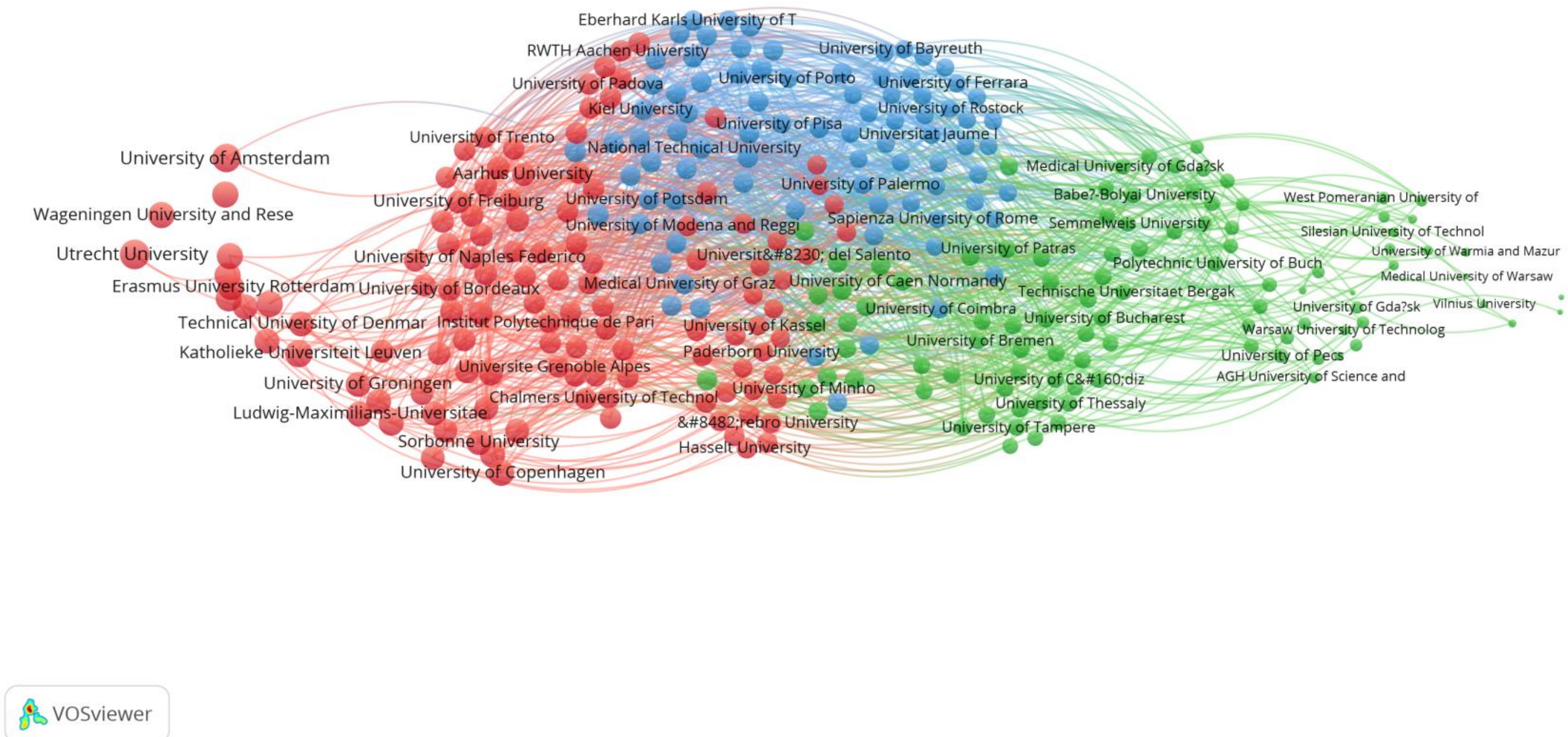


Fig. 1: 302 European universities: three clusters distinguished by the Louvain algorithm; modularity Q = 0.160; Sizes of nodes areproportional to their z-values; links based on significance levels of the chi-square. Parameters of VOSviewer: scale 1.18; size-variation 0.36; attraction = 4; repulsion = 0.

Table 1 reports the top 50 universities in the EU27 which include also the Politecnico di Bari as one of the best Italian universities in the ranking. Interested readers can find the full list of the universities included in the three clusters in SM 2 reported in appendix.

Universities EU27	***z***	**Rank**
Utrecht University	10.472	1
University of Amsterdam	9.341	2
Katholieke Universiteit Leuven	7.553	3
Vrije Universiteit Amsterdam	7.324	4
Erasmus University Rotterdam	7.217	5
Wageningen University and Research	7.126	6
Radboud University	7.012	7
Delft University of Technology	6.785	8
Université Paris Sciences et Lettres	6.450	9
Leiden University	6.292	10
University of Copenhagen	6.081	11
Ludwig-Maximilians-Universitaet Muenchen	5.912	12
University of Groningen	5.897	13
Karolinska Institutet	5.804	14
Technical University of Denmark	5.615	15
Universite de Paris	5.542	16
Trinity College Dublin, The University of Dublin	4.975	17
Sorbonne University	4.815	18
Technical University of Munich	4.801	19
Eindhoven University of Technology	4.710	20
Universite Paris-Saclay	4.452	21
Heidelberg University	4.023	22
Georg-August-Universitaet Goettingen	4.018	23
University of Muenster	4.005	24
Stockholm University	3.828	25
Ghent University	3.747	26
Maastricht University	3.549	27
University of Bonn	3.429	28
University of Naples Federico II	3.301	29
Universite Toulouse III - Paul Sabatier	3.256	30
University College Cork	3.146	31
University of Freiburg	3.085	32
University of Vienna	3.082	33
University of Bordeaux	2.995	34

Karlsruhe Institute of Technology	2.977	35
University of Gothenburg	2.949	36
University of Catania	2.937	37
Aarhus University	2.885	38
University of Antwerp	2.868	39
University of Montpellier	2.858	40
Universite Catholique de Louvain	2.711	41
Julius-Maximilians-Universitaet Wuerzburg	2.504	42
Aalto University	2.353	43
Politecnico di Bari	2.306	44
Tilburg University	2.261	45
Medical University of Vienna	2.227	46
Vita-Salute San Raffaele University	2.220	47
Politecnico di Milano	2.170	48
Johannes Gutenberg University Mainz	2.088	49
University of Trento	2.072	50

Table 1: Top 50 universities in the EU27 ("all sciences," fractionally counted) on the basis of LR 2021. Italian universities are reported in bold.

5.2. ***Clusters of Italian Universities***

According to the Italian Ministry of Universities and Research (MUR), the Italian university system is composed of 97 universities, of which 67 are state universities, 19 legally recognized non-state universities and 11 legally recognized non-state telematics universities.

The state universities are further distinguished into: 56 universities, 3 polytechnics, 6 schools of advanced studies, 2 universities for foreigners.

Forty-two of the Italian universities are included in the Leiden Ranking 2021:

- 3 state polytechnics (Politecnico di Milano, Politecnico di Torino, Politecnico di Bari);
- 2 non-state universities (Vita-Salute San Raffaele University, Università Cattolica del Sacro Cuore);
- 37 State Universities.

Fig. 2 shows the two clusters on the Italian universities obtained by applying the Blondel *et al.* (2008) approach, with the following parameters: NCl=2; Modularity Q = 0.081. Cluster 1 includes 22 universities and Cluster 2 contains 20 universities. Table 2 shows the list of the universities included in the two clusters.

Fig. 2: 42 Italian universities in LR, 2 clusters, Q = 0.081; NCl = 2

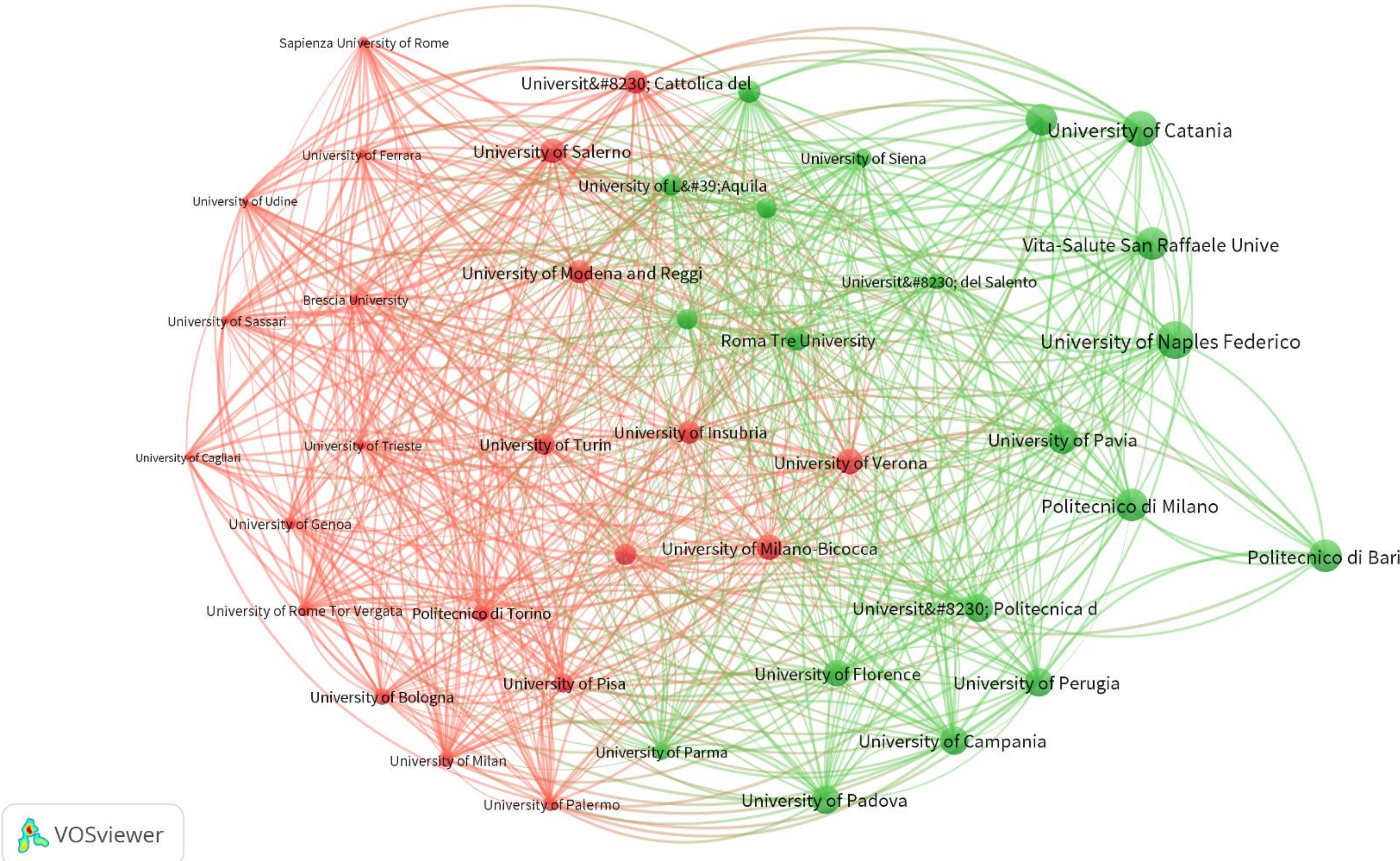
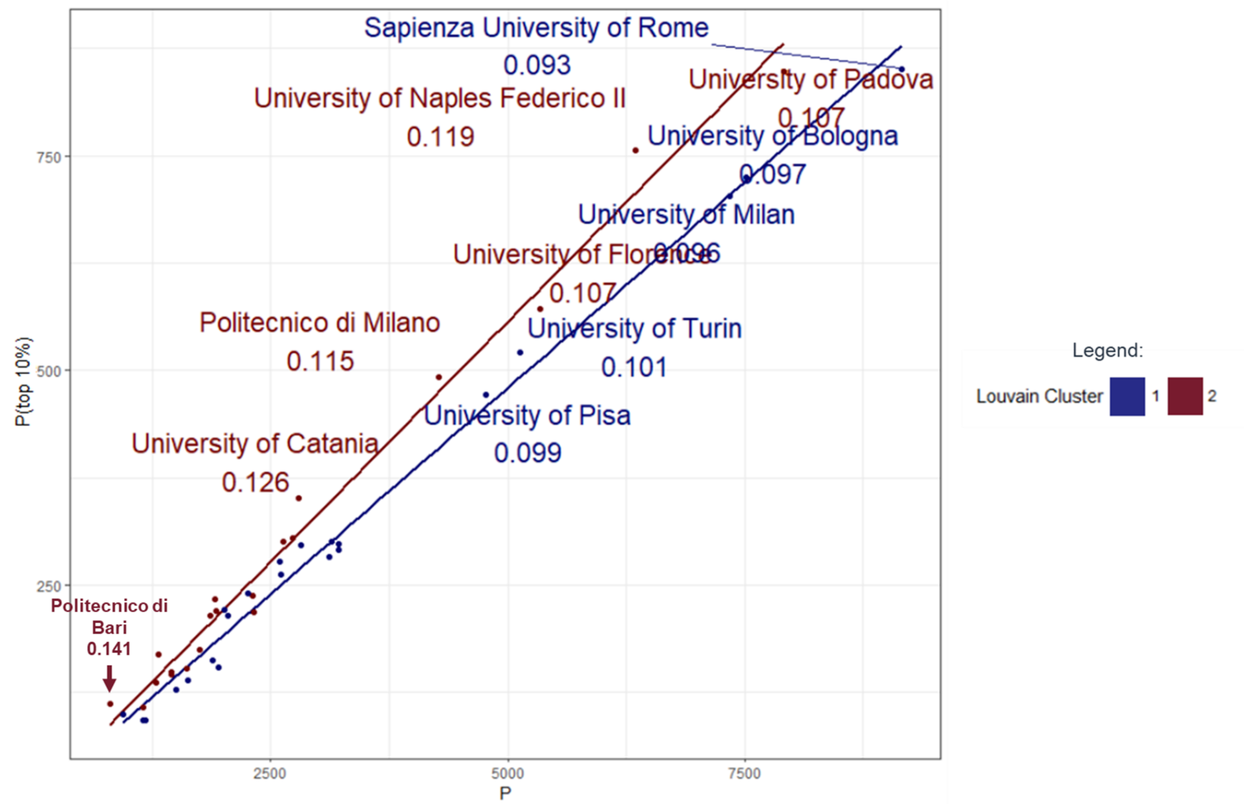


Table 2. List of Italian universities in the 2 clusters.
The Italian universities in the LR 2021 are 42. Cluster 1 contains 22 universities and Cluster 2 contains 20 universities.

Cluster 1	***z* Score**	**Cluster 2**	***z* Score**
University of Verona	0,914	University of Naples Federico II	3,301
University of Milano-Bicocca	0,815	University of Catania	2,937
University of Salerno	0,654	**Politecnico di Bari**	2,306
Università Cattolica del Sacro Cuore	0,612	Vita-Salute San Raffaele University	2,22
University of Modena and Reggio Emilia	0,522	Politecnico di Milano	2,17
University of Insubria	0,424	University of Trento	2,072
University of Turin	0,168	University of Pavia	1,707
University of Bari Aldo Moro	0,12	Università Politecnica delle Marche	1,431
University of Pisa	-0,164	University of Padova	1,423
Politecnico di Torino	-0,539	University of Perugia	1,409
University of Bologna	-0,623	University of Campania Luigi Vanvitelli	1,365
University of Milan	-0,823	University of Florence	1,171
University of Palermo	-0,968	Roma Tre University	0,494
University of Rome Tor Vergata	-1,238	University of Messina	0,335

University of Genoa	-1,256	University of L'Aquila	0,177
University of Trieste	-1,37	Gabriele D'Annunzio University	0,089
University of Ferrara	-1,53	University of Calabria	0
Brescia University	-1,531	University of Siena	-0,486
Sapienza University of Rome	-1,629	Università del Salento	-0,578
University of Sassari	-1,766	University of Parma	-0,581
University of Udine	-1,893		
University of Cagliari	-2,433		

We did some descriptive analysis to try to characterize and differentiate the two groups of Italian universities obtained by applying the Louvain clustering methodology. We analyzed "P(top 10%)"- the number of university publications that, compared to other publications in the same field and the same year, belong to the top 10% most frequently cited – as a function of the total number of publications (P) of a university.



Fig, 3. Number of publications in the top 10% -P(top 10%)- as a function of the total number of publications (P). The Blu line is the regression line of Louvain cluster 1, and the red line is the regression line of Louvain cluster 2).

Fig. 3 shows the linear regressions over the two clusters of Italian universities obtained with the Louvain method. The size of the point is proportional to the PP top 10% and the number reported close to the point is the value of the PP top 10%. It appears that the universities belonging to the cluster two have a higher percentage of papers in the top 10% of most cited works: the regression line of the cluster 2 dominates –*i.e.* is over- the regression line of the cluster one. Using Discriminant Analysis, the difference is statistically significant (Wilks Lambda = .631; p <.010). For example, the Politecnico di Bari is located at the beginning of the regression line of cluster 2, meaning that it has a small number of of publications (P), but it has the highest value of PP top 10%.

We pursued the analysis for the Italian universities by fields of science: Biomedical and health sciences (BIO), Life and earth sciences (LIFE), iii) Mathematics and computer science (MAT), Physical sciences and engineering (FIS); and Social sciences and humanities (SSH).

Table 3 shows the number of P(top 10%) in each of the five fields distinguished in LR (BIO, LIFE, MAT, FIS and SSH) divided by the total number of P(top 10%) in All sciences. For example, considering the Politecnico di Bari, the P(top 10%) in SSH is around 19% of the total P(top 10%), the P(top 10%) in FIS is around 50% of the total P(top 10%), the P(top 10%) in MAT is around 22% of the total P(top 10%), the P(top 10%) in LIFE is around 6% of the total P(top 10%) and the P(top 10%) in BIO is around 2% of the total P(top 10%).

Table 3. Disciplinary composition of Italian universities' P(top 10%) (the number of university publications that, compared to other publications in the same field and the same year, belong to the top 10% most frequently cited). The table reports the number of P(top 10%) in each FOS (BIO, LIFE, MAT, FIS and SSH) over the total number of P(top 10%) in All sciences.

University	**BIO**	**LIFE**	**MAT**	**FIS**	**SSH**
Brescia University	0.605	0.073	0.100	0.180	0.041
Gabriele D'Annunzio University	0.695	0.100	0.028	0.111	0.067
Politecnico di Bari	**0.025**	**0.062**	**0.225**	**0.502**	**0.186**
Politecnico di Milano	0.068	0.076	0.247	0.522	0.086
Politecnico di Torino	0.080	0.076	0.233	0.571	0.040
Roma Tre University	0.084	0.196	0.225	0.357	0.138
Sapienza University of Rome	0.484	0.112	0.100	0.229	0.075
Università Cattolica del Sacro Cuore	0.726	0.104	0.039	0.023	0.107
Università del Salento	0.155	0.170	0.157	0.424	0.094
Università Politecnica delle Marche	0.390	0.264	0.109	0.193	0.045
University of Bari Aldo Moro	0.481	0.255	0.049	0.166	0.049
University of Bologna	0.450	0.168	0.094	0.224	0.064
University of Cagliari	0.362	0.154	0.236	0.203	0.045
University of Calabria	0.189	0.145	0.234	0.377	0.055
University of Campania “Luigi Vanvitelli”	0.654	0.085	0.053	0.137	0.071

University of Catania	0.461	0.193	0.125	0.183	0.038
University of Ferrara	0.584	0.119	0.080	0.183	0.034
University of Florence	0.469	0.184	0.081	0.199	0.067
University of Genoa	0.544	0.108	0.104	0.195	0.050
University of Insubria	0.566	0.180	0.093	0.126	0.035
University of L'Aquila	0.324	0.099	0.220	0.335	0.022
University of Messina	0.501	0.199	0.032	0.230	0.038
University of Milan	0.657	0.143	0.046	0.112	0.043
University of Milano-Bicocca	0.343	0.146	0.187	0.209	0.115
University of Modena and Reggio Emilia	0.576	0.143	0.088	0.146	0.047
University of Naples Federico II	0.353	0.191	0.107	0.303	0.047
University of Padova	0.421	0.160	0.118	0.214	0.087
University of Palermo	0.294	0.200	0.087	0.370	0.050
University of Parma	0.457	0.188	0.101	0.198	0.057
University of Pavia	0.468	0.138	0.110	0.246	0.038
University of Perugia	0.410	0.162	0.131	0.276	0.022
University of Pisa	0.437	0.172	0.150	0.208	0.032
University of Rome Tor Vergata	0.482	0.057	0.126	0.279	0.056
University of Salerno	0.193	0.093	0.197	0.447	0.070
University of Sassari	0.374	0.329	0.064	0.167	0.066
University of Siena	0.637	0.144	0.078	0.072	0.069
University of Trento	0.191	0.117	0.217	0.318	0.157
University of Trieste	0.408	0.115	0.064	0.380	0.033
University of Turin	0.487	0.205	0.042	0.171	0.096
University of Udine	0.340	0.214	0.133	0.240	0.073
University of Verona	0.697	0.095	0.093	0.035	0.080
Vita-Salute San Raffaele University	0.962	0.014	0.003	0.002	0.018

Fig. 4 shows the percentage of publications in the top 10% -PP(top10%)- as a function of the total number of publications (P). The universities belonging to the Louvain cluster 2 have higher values of the percentage of publications in the top 10% -PP(top10%). It seems that the universities in the Louvain cluster 2 are more oriented towards “quality”- as proxied by PP(top10%) than towards “quantity” –as proxied by P. The top-ranked Politecnico di Bari appears in the North-West of the figure, as an outlier. It is characterized by the highest percentage of publications in the top 10%

(PP(top10%)=14%) and by a relatively small total number of publications (P). We will further investigate the Politecnico di Bari in the next section.

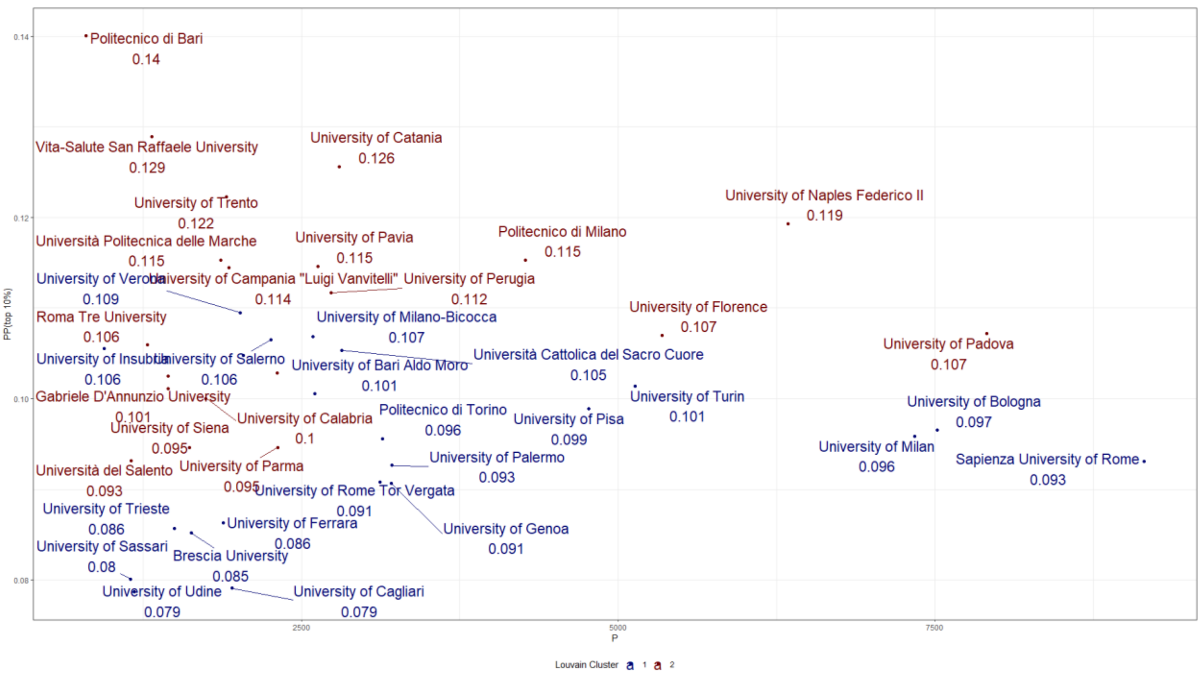


Fig. 4. Percentage of publications in the top 10 PP(top10%) as a function of the total number of publications P (all fields). Universities in blue belong to the Louvain cluster 1, those in red belong to the Louvain cluster 2.

Table 4 shows the *z* scores of all Italian universities, for “All fields” together and the five main Field of Science (FOS) BIO, LIFE, MAT, FIS and SSH compared with their rank for the value of PP(top 10%) in LR. In Table 5, Politecnico di Bari emerges again as the first in terms of PP(top 10%), with a value of 14, while Vita-Salute San Raffaele University is the second with a PP(top 10%) of 12.9. Interestingly, these two universities greatly differ in terms of their disciplinary composition. Vita-Salute San Raffaele University is mostly specialized in BIO with a *z* value of 2.291. The z score of Politecnico di Bari in Social Science and Humanities (SSH) is very high (2.699) compared to its *z* score in “All fields” (2.306) and other major fields, as FIS (1.413).

We calculated Spearman correlations between LR rank according to PP(top 10%) and our z scores "All fields" and Fields of Science (FOC): BIO, LIFE, MAT, FIS and SSH. We obtain a Spearman correlation of -0.99 of LR rank with z "All fields" (4th column of Table 5), of -0.68 with z BIO (5th column of Table 5), of -0.71 with z LIFE (6th column of Table 5), of -0.56 with z MAT (7th column of Table 5), of -0.57 with z FIS (8th column of Table 5) and -0.46 with z SSH (9th column of Table 5). Interestingly, the two fields most correlated to the LR ranks are BIO and LIFE.

Table 4: Comparison of our z scores of the Italian universities with LR for PP(top 10%): All fields and five main Fields of Science (BIO, LIFE, MAT, FIS and SSH)

LR	PP(top 10%)	University	All Fields	BIO	LIFE	MAT	FIS	SSH
1	14	Politecnico di Bari	2.306	-0.378	-0.321	0.857	1.413	2.699
2	12.9	Vita-Salute San Raffaele University	2.22	2.291	0.529	0	0	-0.546
3	12.6	University of Catania	2.937	2.292	1.691	1.254	0.393	0.027
4	12.2	University of Trento	2.072	1.276	1.053	0.858	0.894	0.693
5	11.9	University of Naples Federico II	3.301	1.117	1.45	1.033	2.712	1.194
6	11.5	Politecnico di Milano	2.17	-0.698	0.892	1.076	1.927	0.754
7	11.5	Università Politecnica delle Marche	1.431	0.505	2.02	0.172	0.479	-0.745
8	11.5	University of Pavia	1.707	1.092	0.945	0.986	0.892	-1.129
9	11.4	University of Campania Luigi Vanvitelli	1.365	1.107	0.564	-0.309	0.659	0.606
10	11.2	University of Perugia	1.409	1.348	-0.582	1.379	1.006	-1.557
11	10.9	University of Verona	0.914	0.527	1.082	1.058	-0.501	0.03
12	10.7	University of Padova	1.423	1.051	0.866	1.15	0.099	0.063
13	10.7	University of Florence	1.171	0.907	1.118	-0.265	0.379	-0.242
14	10.7	University of Milano-Bicocca	0.815	-0.383	0.994	2.191	0.165	-0.974
15	10.6	University of Salerno	0.654	-0.324	0.233	0.598	0.641	0.301
16	10.6	Roma Tre University	0.494	-0.316	-0.038	0.569	0.386	0.241
17	10.6	University of Insubria	0.424	0.647	0.572	-0.224	-0.752	-0.063
18	10.5	Università Cattolica del Sacro Cuore	0.612	0.719	0.668	0.44	-0.394	-0.619
19	10.5	University of Modena and Reggio Emilia	0.522	1.337	0.965	-0.438	-1.517	-0.696
20	10.3	University of Messina	0.335	0.08	1.358	-2.017	0	0.166
21	10.2	University of L'Aquila	0.177	-0.765	-0.31	0.422	1.164	-0.794

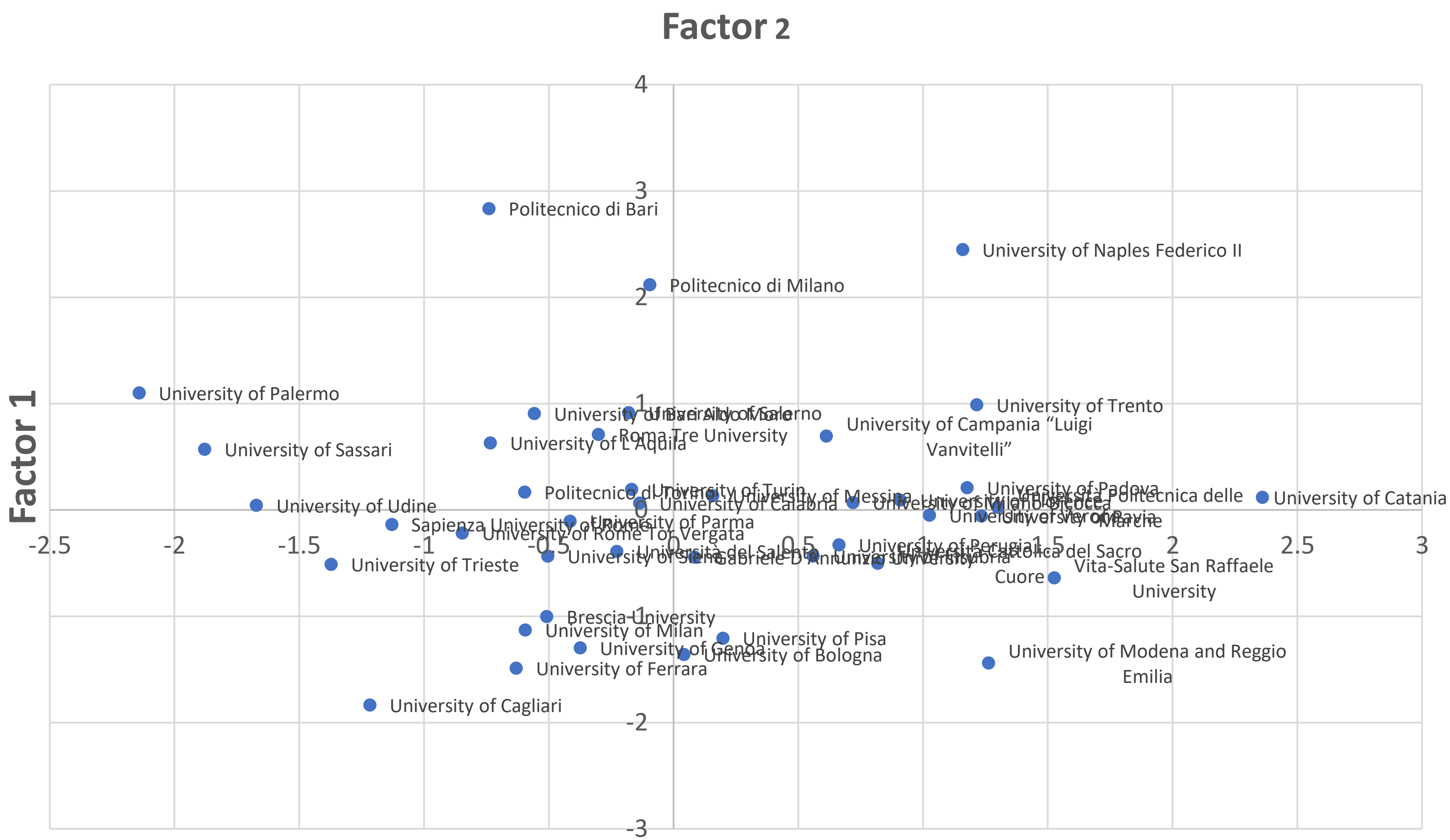
22	10.1	University of Turin	0.168	0.346	0.439	-1.813	0.398	0
23	10.1	Gabriele D'Annunzio University	0.089	0.357	0.16	-0.525	0.573	-1.432
24	10.1	University of Bari Aldo Moro	0.12	-0.721	0.558	-0.775	1.218	-0.229
25	10	University of Calabria	0	0	-0.326	0.765	-0.386	0.045
26	9.9	University of Pisa	-0.164	0.607	-0.583	0.571	-0.64	-1.527
27	9.7	University of Bologna	-0.623	0	0.161	-0.462	-0.293	-2.237
28	9.6	University of Milan	-0.823	0.469	-1.414	-0.468	-1.253	-0.609
29	9.6	Politecnico di Torino	-0.539	-0.277	-0.11	-0.923	0	0
30	9.5	University of Parma	-0.581	-0.473	-0.199	0.292	-0.791	0.052
31	9.5	University of Siena	-0.486	0.216	-0.712	-0.675	-0.464	-0.352
32	9.3	Università del Salento	-0.578	0.467	-0.433	-0.711	-0.21	-0.448
33	9.3	Sapienza University of Rome	-1.629	-1.492	0.072	-0.928	-0.105	-0.885
34	9.3	University of Palermo	-0.968	-2.054	-1.159	-0.198	1.318	-0.501
35	9.1	University of Rome Tor Vergata	-1.238	-0.756	-0.225	-0.845	-0.651	-0.186
36	9.1	University of Genoa	-1.256	-0.095	-0.572	-0.307	-1.649	-0.8
37	8.6	University of Ferrara	-1.53	-0.148	-1.139	-0.055	-1.639	-1.15
38	8.6	University of Trieste	-1.37	-1.173	-0.618	-1.294	0.295	-1.674
39	8.5	Brescia University	-1.531	-0.589	0.218	-1.343	-1.288	-0.738
40	8	University of Sassari	-1.766	-1.932	-1.279	0.524	0.269	-0.406
41	7.9	University of Cagliari	-2.433	-1.938	-1.012	1.14	-1.852	-2.419
42	7.9	University of Udine	-1.893	-1.667	-0.449	-1.472	-0.185	-0.446

To better understand the relationships among the z scores by FOS we run a principal components analysis among them. The rotation method applied is Varimax with Kaiser Normalization. Rotation converged in 3 iterations. The two main factors obtained, Factor 1 and Factor 2, which explain 64.4% of the variance, are illustrated in Fig. 5.

SM 3, in appendix, lists Italian universities according to the two main factors of the Principal component analysis illustrated in Fig. 5. Fig. 6 shows the components of the two main factors that are reported in Table 5. We observe that Factor 1 is mostly characterized by the BIO and LIFE fields of science, while Factor 2 is mainly characterized by FIS and SSH that are highly correlated. The surprisingly high correlation between FIS and SSH may be due to the fact that FIS includes physical sciences and engineering and that SSH includes publications indexed in Web of Science in social science and humanities that could be more quantitative. This might explain why polytechniques are located on the Factor 2 axe as shown in Fig. 5.

Fig. 7 shows a geographical map of the localization of the Italian universities included in the LR. Contrary to common wisdom, it appears that the two Lovain groups detected (group 1 and group 2) are all spread across the country, having many highly performing universities in the South of Italy and showing that in the same geographical areas coexist universities from the two groups (for instance in the city of Bari there is the Politecnico di Bari that belongs to Group 2 and University of Bari Aldo Moro that belongs to Group 1).

Fig. 5. Principal components analysis on the z scores by FOS.



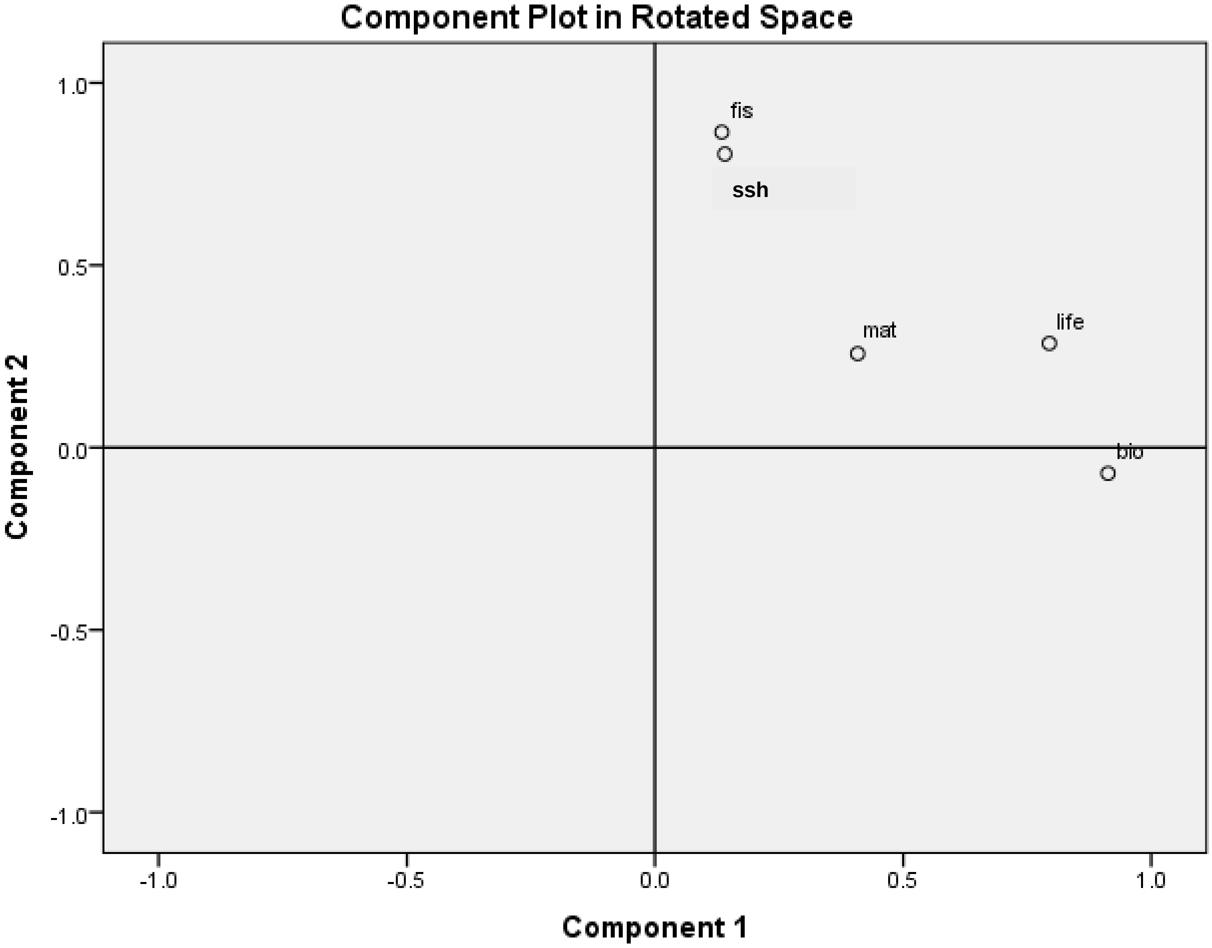


Fig. 6. Principal components analysis on the z scores by FOS: Component Plot.

	Component	
	1	2
BIO	**.912**	
LIFE	**.794**	.286
MAT	.408	.258
FIS	.134	**.865**
SSH	.141	**.805**

Table 5. Principal components analysis on the z scores by FOS.

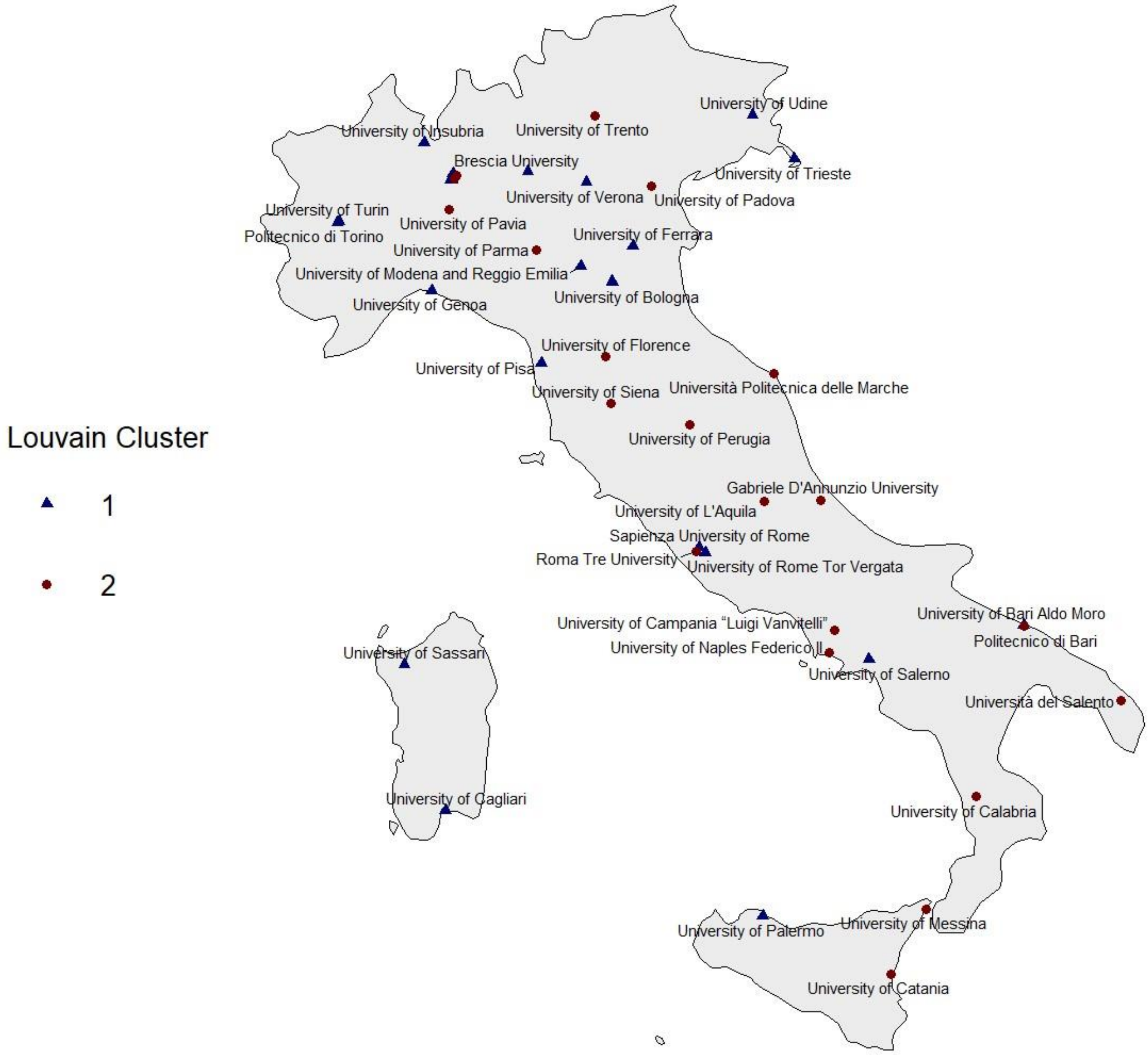


Fig. 7: Geographical map of the 42 Italian universities in the LR. 2 Louvain clusters (Q = 0.081; NCl = 2).

*5.3. **The case of Politecnico di Bari***

All the analyses reported in the previous sections point to the Politecnico di Bari (Polytechnic of Bari) as an outlier. We analyzed this is in more detail. According to the Leiden Ranking, Politecnico di Bari is the top-ranked Italian university and the 37th in Europe for PP top 10%, with a total of 800 publications (fractional count) and 14% of them in the top 10%, considering all fields. Politecnico di Bari was founded in 1990. It is the only polytechnic in the south of Italy and also the youngest polytechnic in Italy. Its disciplinary specialization is in the fields of architecture, engineering, and industrial design.

In the category Social Science and Humanities (SSH), the Politecnico di Bari ranks first not only in Italy but also in Europe and it is ranked second worldwide. Full details of the Politecnico di Bari's results in Leiden Rankings are shown in the Table 6 with the values both with and without fractional counts.

Field	**P**	**P_top10**	**PP_top10**	**P_Fract**	**P_top10_fract**	**PP_top10_fract**
All sciences	2051	319	15.572%	800	112	14.000%
BIO	72	4	5.721%	37	3	7.595%
LIFE	159	13	8.474%	82	7	8.565%
MAT	332	48	14.412%	195	25	12.894%
FIS	1388	222	15.962%	422	56	13.331%
SSH	101	32	**32.244%**	64	21	**32.385%**

Table 6 Research output by fields of science of the Politecnico di Bari in LR 2021

We analyzed the Politecnico di Bari's publications in social science and humanities (SSH) to try to understand the causes behind its Italian, European and even worldwide excellence in this field. The publications identified for these analyzes were obtained through the process detailed in SM 4, in appendix.

Fig. 8 shows the co-authorship analysis performed on the SSH publications of Politecnico di Bari using VosViewer.

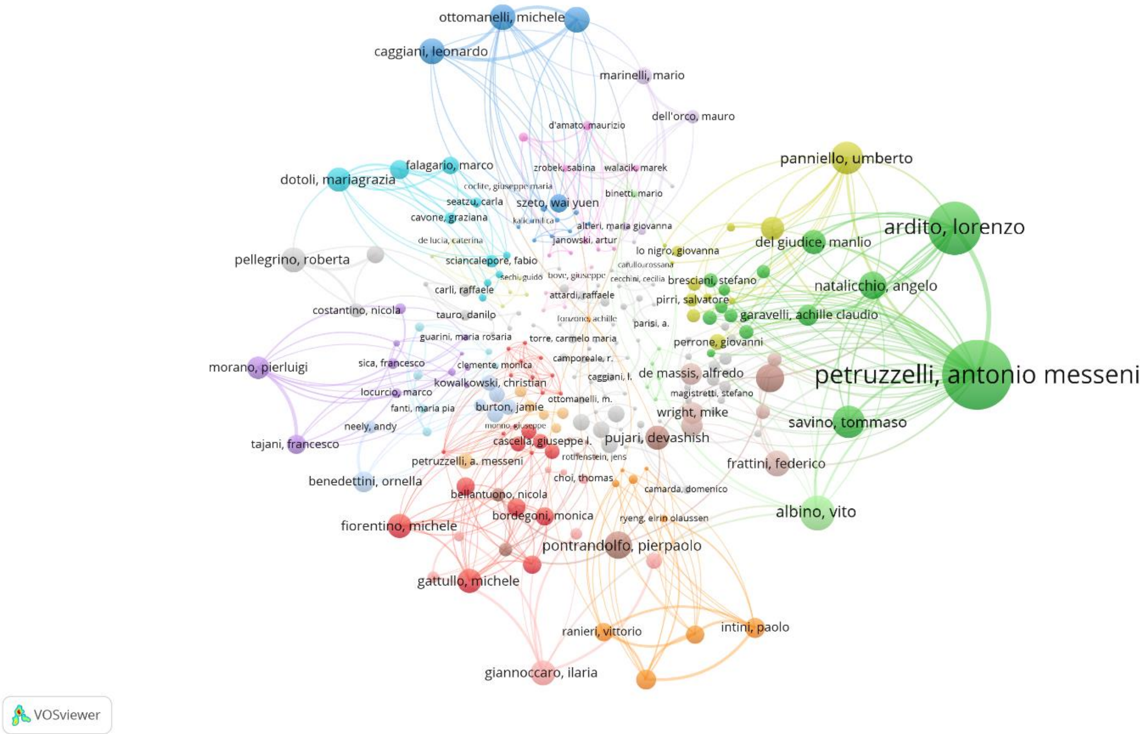


Fig. 8 Co-authorships map of the publications in SSH of Politecnico di Bari

Fig. 8 shows that a prominent role in the Politecnico di Bari's SSH publication network is played by the green sub-network, located in the eastern part of the graph. This sub-network is driven by a young and promising scholar, Antonio Messeni Petruzzelli and another emerging star that is Lorenzo Ardito. Antonio Messeni Petruzzelli is full professor of Management Engineering. He was born on 10 February 1980 in Bari and did his undergraduate studies and post-graduate studies, including his PhD, at the Politecnico di Bari. Lorenzo Ardito is assistant professor of Management Engineering at the Politecnico di Bari, where he did his studies, including his PhD under the supervision of Antonio Messeni Petruzzelli.

Fig. 9 shows the three fields plot performed using the R package bibliometrix (Aria and Cuccurullo, 2017). It shows the relationships of institutions, authors, and authors keywords.

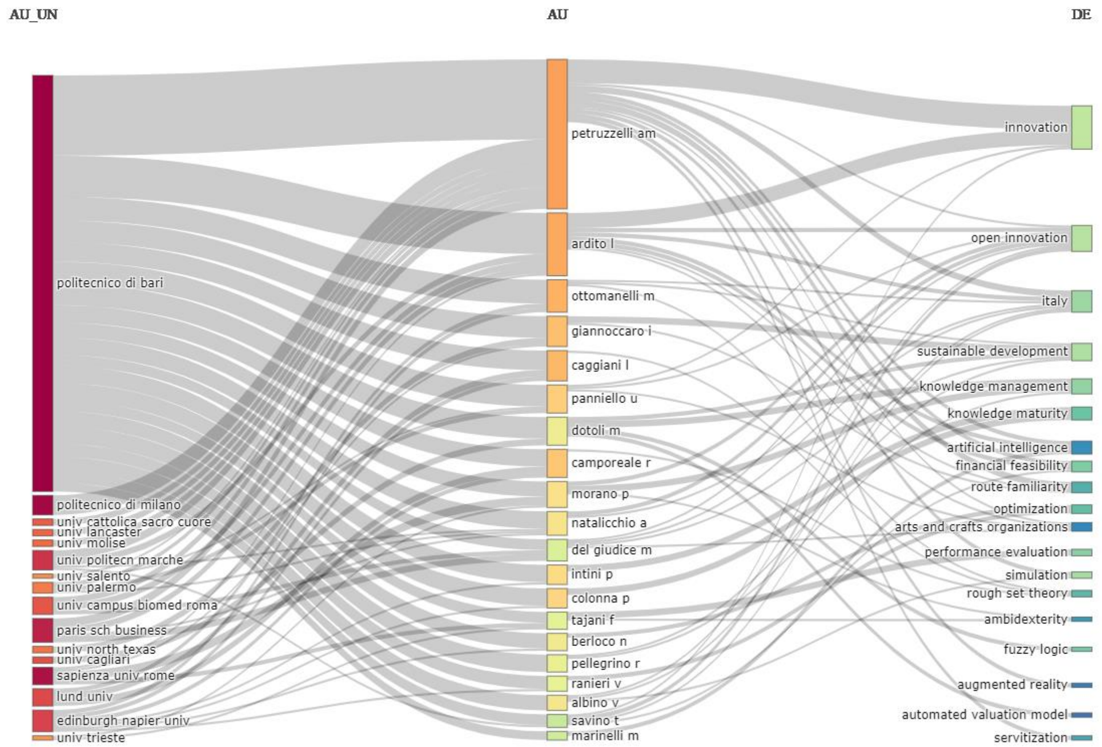


Fig. 9 Three fields plot. From the left to the right: Institutions (AU_UN), authors (AU) and authors keywords (DE)

The inspection of Fig. 9 shows that most of the scientific production in SSH of Politecnico di Bari is driven by Antonio Messeni Petruzzelli and by Lorenzo Ardito already noted in Fig. 8. In addition we have also information about the main keywords of the scientific production of Politecnico di Bari which include innovation, open innovation, Italy, sustainable development, knowledge management, knowledge maturity, artificial intelligence, financial feasibility, route familiarity, optimization, art and craft organizations, performance evaluation, simulation, rough set theory, fuzzy logic, augmented reality, automated valuation model, “servitization”.

We asked to a few members of the Politecnico di Bari’s team, namely to Vito Albino the historical leader of the group, Antonio Messeni Petruzzelli and Rosa Maria Dangelico to give us the list of topics in which they are active and to describe us the strategy they used in the organization of their research team.

We observe an important overlap between the authors keywords shown in Fig. 7 which come from our bibliometric analysis based on the publications retrieved from WoS and the main topics in which the team declared to being active in, which are the following:

- knowledge search and recombination;
- R&D alliance;
- industrial symbiosis;
- crowdfunding;
- patent analysis;
- digital innovation;
- sustainability.

Politecnico di Bari's team implemented the following strategies: (i) orientation to social challenges; (ii) creation of national and international research networks; (iii) thematic specialization among members of the team and (iv) valorization of team members with a high propulsive thrust.

6. Discussion and policy implications

Understanding, modeling and evaluating the scientific performance of universities is a complex activity. Scientific performance is influenced by various factors and by the political-institutional context. Without claiming to be exhaustive, the available funding and infrastructures, the complementarity and substitutability of other activities carried out -as teaching and the so-called "third mission"-, the contribution provided by the technical and administrative staff of the university (Avenali et al., 2022), are some of the factors that influence the performance of research, in addition to scientific merit and individual capabilities of academic staff.

The Italian university system is characterized by an “endemic” public underfunding, and it is an “undifferentiated” system in which all academics have to do both research and teaching. Universities then, having similar institutional incentives, show a certain degree of isomorphism (DiMaggio and Powell, 1983) which makes it difficult to discriminate between them. In this context, existing university rankings, including LR, can hardly help to understand what lies behind the rank and indicator numbers reported in the university rankings. Moreover, Bruni et al. (2020) studying the heterogeneity of European universities, including Italian universities, warned against using one-dimensional approaches to the performance of higher education institutions.

Hence, the analyses carried out in this paper using the LR are far from being a comprehensive and complete analysis of Italian universities. This is due, in addition to the theoretical problems that were mentioned above on the complexity of the assessment of research, also to several potential limitations and biases in the analyses performed, which can be attributed to problems with the data and methodologies adopted.

First of all, Leiden Rankings share the main limits of other existing rankings (see e.g. Daraio and Bonaccorsi, 2017 and Fauzi et al. 2020) considering only the publications indexed in the Web of Science database to build the bibliometric indicators proposed. It is well known that the coverage of Web of Science differs by discipline, having social science and humanities less represented in terms of outputs reported in WoS compared to the total production of the considered universities.

Secondly, the coverage of universities in the LR is not complete. Of the 97 Italian universities, only 42 are included in the LR.

Thirdly, the classification by fields of science proposed in the LRs is very aggregated and combines together two quite different disciplinary fields such as physics and engineering. This may imply, for example, that universities with a solid tradition of excellence in physics, such as the University of Rome La Sapienza, whose professor Giorgio Parisi won the Nobel Prize for Physics in 2021, are not well positioned in the LR.

Fourthly, in this paper, we analyzed only a small number of indicators, focusing mainly on the PP(top10%) indicator, i.e., considering the percentages of publications in the top 10%. Although this is a percentile indicator and therefore "size-independent", the analyses we have carried out show that the Politecnico di Bari outperforms all other Italian universities with a PP(top10%) equal to 14% and 112 publications in the top 10% out of a total of 800 publications (all science disciplines together, fractional counting). If we consider larger universities such as Sapienza university, we notice that it shows a PP(top10%) equal to 9.3% and 852 publications in the top 10% our of a total of 9150 publications, much larger than that of Bari Polytechnic. It seems then more difficult for larger universities to have a high PP(top10%). Furthermore, an important role is played by the disciplinary specialization of the university which, if generalist, suffers more from the poor coverage of social science and humanities research outputs in the WoS database.

Finally, the methodology used by Leydesdorff et al. (2019) may be subject to the limits of the use of statistical significance tests in research assessment that are well known in the existing literature (see e.g. Schneider, 2013). However, in this paper we use the statistical information with an exploratory purpose, i.e. for starting an activity of further investigation to analyze more in-depth the identified outlying institutions through case studies. Moreover, to check the robustness of our analyses, we provided several rank correlations among the results obtained by applying the statistical test procedure proposed by Leydesdorff et al. (2019) and the bibliometric indicators reported in the LR.

Taking all of the above into consideration, we looked at research excellence only as proxied by PP(top10%) to provide an analytical window from a national perspective, and find our analysis informative.

The new way of looking and using LR we propose in this paper could be further explored and linked to the "science of science" perspective (Wang and Barabási, 2021) by analyzing the performance of "emerging" groups or scholars within their networks. The creation and assembly mechanisms of groups influence both the structure of collaboration networks - analysed through co-authorship, and the performance of scientific teams (Guimera et al. 2005). As shown by Wuchty et al. (2007), there is an increasing dominance of relatively small groups in the production of knowledge. In all disciplines, both for scientific production (measured in terms of publications) and for innovative production (measured in terms of patents), research is increasingly conducted in teams, which produce more cited and high-impact works than individual works.

Although there is a growing trend (Wu et al., 2019) in all scientific fields towards the presence of large teams, Wu et al. (2019) showed that alongside large teams there are small groups that are responsible for high impact research. Small teams tend to produce disruptive work while large teams are inclined to carry out developing work. According to Wu et al. (2019) one major implication is that both small and large teams are essential to scientific development.

The problem of optimal allocation of research funds to researchers, groups, and departments within a university is a long-standing one to which too few studies have been devoted. Despite some notable exceptions, such as Ioannidis (2011) and Stephan (2012), the problem of how to best fund individuals and groups within universities is still not thoroughly addressed today.

Wu et al. (2019) suggested that research policies should aim to support both large and small teams. Along the same lines, Moed (2017, pp. 150-151) proposed a method of funding basic research which would be expected to cushion the Matthew effect implied when funding is given on the basis of previous performance. He proposed to focus on “emerging groups” rather than on the total academic staff of universities. An emerging group is considered to be a small research group that is expected to have great research potential. The director of this group should normally be a young “rising star” with a promising research agenda. The assessment of these emerging groups should be based on quantitative minimum standards in terms of bibliometric indicators and peer review. The main idea behind this type of research policy, based on the awarding of emerging groups, is to support and develop good research practices. As shown by Daraio and Vaccari (2020), reflection on good research practices and the role of researchers in them provides important information for both individual development and self-assessment and improvement. The results of our approach might be useful in helping to identify productive and emerging researchers and groups.

7. Concluding remarks

In this work we proposed a new use of Leiden rankings as a source of heuristic information to be deepened with case studies of institutions that are particularly performing (outlier). This new way of using the information of the LR started by applying statistical inference, mapping and clustering of Italian universities within Europe. After that, we looked for prominent teams that appeared to be present in one of the outlier that we identified in the previous analysis, namely Politecnico di Bari. In particular, we identified an “emerging team” lead by a young and promising scholar.

Thanks to the elaborations carried out on the LR and the case study on the Politecnico di Bari, we were able to identify a small team lead by a young and promising scholar who pushes the highest scientific results of the whole university. This shows that the organization of a research team and the choice of the topics of research and of journal outlets can affect the positioning of a university in the LR. Furthermore, the analyses carried out showed that the disciplinary composition on the research of a university can impact the overall position of that university in LR. This is a topic that should be further investigated.

We aim to contribute to developing the ongoing research activity on rankings, their limits and their potential, linking this research with the research line on teams and that on good research practices.

Finally, we discussed the implications of the results in terms of research policies, in support of "emerging teams", as suggested by Moed (2017), as an alternative to other research support policies, to be adopted to encourage virtuous behavior and best practices in research activity.

Acknowledgements

We thank Vito Albino, Alessandro Avenali, Rosa Maria Dangelico and Antonio Messeni Petruzzelli for the comments provided during the writing of the work.

Website References

https://www.leidenranking.com, last access 18/11/2021, 15.30

https://www.webofscience.com/wos/woscc/basic-search, last access 18/11/2021, 14.00

https://www.leidenranking.com/information/fields, last access 18/11/2021, 13.30

https://www.miur.gov.it/web/guest/istituzioni-universitarie-accreditate, last access 28/12/2021, 15.30

8. References

Abramo, G., D'Angelo, A. C., Grilli, L. (2016). From rankings to funnel plots: The question of accounting for uncertainty when assessing university research performance. *Journal of Informetrics*, 10(3), 854-862.

Aria, M. Cuccurullo, C. (2017) bibliometrix: An R-tool for comprehensive science mapping analysis, *Journal of Informetrics*, 11(4), pp 959-975.

Avenali, A., Daraio C., Wolszczak-Derlacz J. (2022), Determinants of the incidence of non-academic staff in European and US HEIs, *Higher Education,* 1-29.

Blondel, V. D., Guillaume, J. L., Lambiotte, R., Lefebvre, E. (2008). Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10), P10008.

Bornmann, L., Leydesdorff, L., Mutz, R. (2013). The use of percentiles and percentile rank classes in the analysis of bibliometric data: Opportunities and limits. *Journal of informetrics*, 7(1), 158-165.

Bruni, R., Catalano, G., Daraio, C., Gregori, M., Moed, H. F. (2020). Studying the heterogeneity of European higher education institutions. *Scientometrics*, 125(2), 1117-1144.

Calero-Medina, C., Noyons, E., Visser, M., De Bruin, R. (2020). Delineating Organizations at CWTS—A story of many pathways. In *Evaluative Informetrics: The Art of Metrics-Based Research Assessment: Festschrift in Honour of Henk F. Moed*, edited by Daraio C. and Glanzel W., pp. 163-177, Springer Nature.

Cohen, J. (1977). *Statistical Power Analysis for the Behavioral Sciences*, Orlando, FL, etc.: Academic Press.

Colliander, C., Ahlgren, P. (2011). The effects and their stability of field normalization baseline on relative performance with respect to citation impact: A case study of 20 natural science departments. *Journal of Informetrics*, 5(1), 101-113.

Dahler-Larsen, P. (2011). *The evaluation society*. Stanford, CA: Stanford University Press.

Daraio C., Bonaccorsi A., Simar L. (2015), Rankings and University Performance: a Conditional Multidimensional Approach, *European Journal of Operational Research*, 244, 918-930.

Daraio, C., Bonaccorsi, A. (2017). Beyond university rankings? Generating new indicators on universities by linking data in open platforms. *Journal of the Association for Information Science and Technology*, 68(2), 508-529.

Daraio, C., Vaccari, A. (2020). Using normative ethics for building a good evaluation of research practices: towards the assessment of researcher's virtues. *Scientometrics*, 125(2), 1053-1075.

DiMaggio, P. J., Powell, W. W. (1983). The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. *American Sociological Review*, 147-160.

Fauzi, M. A., Tan, C. N. L., Daud, M. and Awalludin, M. M. N. (2020), “University rankings: A review of methodological flaws”, *Issues in Educational Research*.

Guimera R, Uzzi B, Spiro J, Amaral LAN (2005) Team assembly mechanisms determine collaboration network structure and team performance. *Science* 308(5722):697–702.

Härdle, W., Simar, L. (2019). *Applied multivariate statistical analysis* (pp. 431-442). Springer International Publishing.

Hazelkorn E., Mihut G. (2021) eds, *Research Handbook on University Rankings: Theory, Methodology, Influence and Impact*, Edward Elgar, Cheltenam (UK).

Ioannidis, J. (2011). Fund people not projects. *Nature*, 477(7366), 529-531.

Law 30 of December 2010, n. 240, Norme in materia di organizzazione delle università, di personale accademico e reclutamento, nonchè delega al Governo per incentivare la qualità e l'efficienza del sistema universitario, Gazzetta Ufficiale n. 10 del 14 gennaio 2011.

Leydesdorff, L., Bornmann, L. (2012). Testing differences statistically with the Leiden ranking. *Scientometrics*, 92(3), 781-783.

Leydesdorff, L., Wagner C. S., and Zhang L. (2021), Are University Rankings Statistically Significant? A Comparison among Chinese Universities and with the USA. *Journal of Data and Information Science*, vol.6, no.2, 2021, pp.67-95.

Moed, H. F. (2017). A critical comparative analysis of five world university rankings. *Scientometrics*, 110(2), 967-990.

Moed, H. F. (2017). *Applied evaluative informetrics*. Berlin: Springer International Publishing.

Moed, H. F., Halevi, G. (2015). Multidimensional assessment of scholarly research impact. *Journal of the Association for Information Science and Technology*, 66(10), 1988-2002.

Olcay, G. A. and Bulu, M. (2017), “Is measuring the knowledge creation of universities possible?: A review of university rankings”, *Technological Forecasting and Social Change*, Vol. 123, pp. 153-160.

Pearson K. (1901). LIII. On lines and planes of closest fit to systems of points in space. *The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science*, 2(11), 559–572.

Schneider, J. W. (2013). Caveats for using statistical significance tests in research assessments. *Journal of Informetrics*, 7(1), 50-62.

Stephan, P. (2012). *How economics shapes science*. Harvard University Press.

Strotmann, A., Zhao, D. (2015). An 80/20 data quality law for professional scientometrics? In ISSI 2015 Proceedings.

Van Eck, N. J., Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics*, 84(2), 523-538.

Vernon, M. M., Balas, E. A. and Momani, S. (2018), "Are university rankings useful to improve research? A systematic review", *PloS one*, Vol. 13, No. 3.

Waltman, L., Calero-Medina, C., Kosten, J., Noyons, E. C., Tijssen, R. J., van Eck, N. J., ... Wouters, P. (2012). The Leiden Ranking 2011/2012: Data collection, indicators, and interpretation. *Journal of the American society for information science and technology*, 63(12), 2419-2432.

Wang, D., Barabási, A. L. (2021). *The science of science*. Cambridge University Press.

Wu, L., Wang, D., Evans, J. A. (2019). Large teams develop and small teams disrupt science and technology. *Nature*, 566(7744), 378-382.

Wuchty S, Jones BF, Uzzi B (2007) The increasing dominance of teams in production of knowledge. *Science*, 316(5827):1036–1039.

Supplementary Materials

SM 1. Data extraction and routine to generate the empirical results

For our purposes, we reorganize the file so that the fields "university," "country," "field," "period" (publication years), "fractional" (fractional or full counting of publications), "P" (number of publications), "P top10%" "PP top10%" and its upper and lower bounds are saved as a comma-separated data file. A dedicated routine (available at http://www.leydesdorff.net/software/leiden) reads this file as input and generates, for each country and the whole set the information relevant for the analysis as detailed below.

The following files are generated by the execution of the routine:

1. A file in the Pajek format with universities as vertices and *z*-values as links insofar as $z < 2.576$ (the cutoff for $p < .01$). This file is named with the country name (e.g., "Germany.net" describing 50 German universities represented in the data). Files are thus generated for the 54 individual countries, and one additional file "world.net" contains the data for all 902 universities.
2. A second file in the Pajek format with similar information, but with **w-** values for the links; in this case, no threshold is set *a priori*. Each of these files has the same name as under 1, but "_w" is added to the country name as a root (e.g., "Germany_w.net").
3. A file in the Pajek format with similar information, but with value 1 for the links between universities with overlapping stability intervals, and 2 for the strong components. Each of these files has the same name as under 1, but "_o" is added to the country name as a root.
4. The *z*-values for testing the universities at the nodes against the 10% value of most highly-cited publications are stored in a file with the same country names, but with the extension

“.vec”. Since most network programs can handle only positive values, negative values of *z* are set equal to zero.

5. The full set of *z* values is retrievable from the original .dbf files as a field and from the Pajek files in the header, indicating the size of each node. Using these files, positive values can be represented in the visualization (using Pajek) by a circle and negative ones by a diamond. A partition file with the extension “.clu” for each country is generated distinguishing between positive and negative values of *z* (using “2” and “1”, respectively.)

The above files are made for all countries under study. The program EU.exe makes additionally a similar file for the EU27. The Pajek (.net) format provides a kind of currency among programs for network analysis and visualization. We store the resulting measures (overlap, *z*, or ***w***) in the edge value between each two universities. Note that these universities are not necessarily related for example by citation or co-authorship. The use of network statistics is in this sense metaphorical. However, both VOSviewer and MDS (e.g., in NetDraw or SPSS) are based on showing structural similarities (in a vector space), in our case similar or different institutional impact performances in terms of statistical and practical significance. In VOSviewer, network links can additionally facilitate the interpretation.

SM 2. List of European universities in the obtained clusters

Cluster 1	**Cluster 2**	**Cluster 3**
Åbo Akademi University	Lodz University of Technology	Aix-Marseille University
Örebro University	Adam Mickiewicz University in Poznań	Aristotle University of Thessaloniki
Aalborg University	AGH University of Science and Technology	Brescia University
Aalto University	Babeș-Bolyai University	Carl von Ossietzky University of Oldenburg
Aarhus University	Brno University of Technology	Claude Bernard Lyon 1 University
Bielefeld University	Budapest University of Technology and Economics	Eberhard Karls University of Tuebingen
Chalmers University of Technology	Carlos III University of Madrid	Friedrich Schiller University Jena
Delft University of Technology	Charles University	Gottfried Wilhelm Leibniz Universitaet Hannover
Dublin City University	Chemnitz University of Technology	Heinrich Heine University Duesseldorf
Eindhoven University of Technology	Comenius University in Bratislava	Justus Liebig University Giessen
Erasmus University Rotterdam	Complutense University of Madrid	Kiel University
Freie Universitaet Berlin	Czech Technical University in Prague	KTH Royal Institute of Technology
Friedrich-Alexander-Universitaet Erlangen-Nuernberg	Czech University of Life Sciences Prague	Leipzig University
Gabriele D'Annunzio University	Democritus University of Thrace	Linkoeping University
Georg-August-Universitaet Goettingen	Eötvös Loránd University	Luleå University of Technology
Ghent University	Gdańsk University of Technology	Martin Luther University Halle-Wittenberg
Goethe University Frankfurt	Jagiellonian University	Medical University of Graz
Graz University of Technology	Maria Curie-Sklodowska University	Medical University of Innsbruck
Gustave Eiffel University	Masaryk University	Miguel Hernández University
Hannover Medical School	Medical University of Lodz	National Technical University of Athens
Hasselt University	Medical University of Białystok	Palacký University Olomouc
Heidelberg University	Medical University of Gdańsk	Philipps-Universitaet Marburg
Humboldt-Universitaet zu Berlin	Medical University of Lublin	Politecnico di Torino
Institut National Polytechnique de Toulouse	Medical University of Silesia	Rey Juan Carlos University
Institut Polytechnique de Paris	Medical University of Warsaw	Roma Tre University

Johannes Gutenberg University Mainz	National and Kapodistrian University of Athens	Ruhr-Universitaet Bochum
Johannes Kepler University Linz	National Distance Education University	Sapienza University of Rome
Julius-Maximilians-Universitaet Wuerzburg	Nicolaus Copernicus University in Toruń	Technische Universitaet Darmstadt
Karlsruhe Institute of Technology	Otto von Guericke University Magdeburg	Technische Universitaet Dresden
Karolinska Institutet	Polytechnic University of Bucharest	Umeå University
Katholieke Universiteit Leuven	Poznan University of Life Sciences	Universidad Autonoma de Madrid
Leiden University	Poznan University of Medical Sciences	Universidade Nova de Lisboa
Ludwig-Maximilians-Universitaet Muenchen	Poznan University of Technology	Università Cattolica del Sacro Cuore
Lund University	Saarland University	Universidad Autónoma de Barcelona
Maastricht University	Semmelweis University	Universitat Jaume I
Medical University of Vienna	Silesian University of Technology	Universitat Rovira i Virgili
National University of Ireland, Galway	Slovak University of Technology in Bratislava	Universite de Lorraine
Paderborn University	Technical University of Madrid	Universite de Nantes
Politecnico di Bari	Technische Universitaet Bergakademie Freiberg	Universite de Versailles Saint-Quentin-en-Yvelines
Politecnico di Milano	Technische Universitaet Braunschweig	University of A Coruña
Pompeu Fabra University	TU Dortmund University	University of Alcalá
Radboud University	Ulm University	University of Aveiro
RWTH Aachen University	Universidad de Almería	University of Bari Aldo Moro
Sorbonne University	Universidad de La Laguna	University of Bayreuth
Stockholm University	Universidade de Lisboa	University of Bologna
Swedish University of Agricultural Sciences	Universidade de Vigo	University of Clermont Auvergne
Technical University of Denmark	Universitat Politecnica de Catalunya, BarcelonaTech	University of Coimbra
Technical University of Munich	Universitat Politecnica de Valencia	University of Eastern Finland
Technische Universitaet Berlin	Universite de Tours	University of Ferrara
Technische Universitaet Kaiserslautern	University of Lodz	University of Genoa
Tilburg University	University of Agriculture in Krakow	University of Granada
Trinity College Dublin, The University of Dublin	University of Alicante	University of Greifswald
Università del Salento	University of Bremen	University of Helsinki
Università Politecnica delle Marche	University of Bucharest	University of Hohenheim
Universitaet Hamburg	University of Cadiz	University of Liege

Universitaet Regensburg	University of Caen Normandy	University of Lille
Universitaet zu Luebeck	University of Cagliari	University of Malaga
Universite Cote d'Azur	University of Cantabria	University of Milan
Universite Catholique de Louvain	University of Castilla-La Mancha	University of Milano-Bicocca
Universite de Paris	University of Chemistry and Technology, Prague	University of Natural Resources and Life Sciences, Vienna
Universite Grenoble Alpes	University of Cordoba	University of Navarra
Universite Libre de Bruxelles	University of Debrecen	University of Oulu
Universite Paris Sciences et Lettres	University of Extremadura	University of Palermo
Universite Paris-Est Creteil Val de Marne	University of Gdansk	University of Pisa
Universite Paris-Saclay	University of Graz	University of Porto
Universite Sorbonne Paris Nord	University of Innsbruck	University of Potsdam
Universite Toulouse III - Paul Sabatier	University of Insubria	University of Rome Tor Vergata
University College Cork	University of Jaen	University of Rostock
University College Dublin	University of Jyvaeskylae	University of Rouen
University of Amsterdam	University of Las Palmas de Gran Canaria	University of Salerno
University of Angers	University of Life Sciences in Lublin	University of Salzburg
University of Antwerp	University of Ljubljana	University of Santiago de Compostela
University of Barcelona	University of Minho	University of Sassari
University of Bonn	University of Murcia	University of Southern Denmark
University of Bordeaux	University of Orleans	University of Tartu
University of Calabria	University of Oviedo	University of the Basque Country
University of Campania	University of Parma	University of Trieste
University of Catania	University of Patras	University of Turin
University of Cologne	University of Pecs	University of Turku
University of Copenhagen	University of Salamanca	University of Udine
University of Crete	University of Seville	University of Verona
University of Cyprus	University of Siena	Vienna University of Technology
University of Duisburg-Essen	University of Silesia in Katowice	Vrije Universiteit Brussel
University of Florence	University of Szeged	

University of Freiburg	University of Tampere	
University of Girona	University of Thessaly	
University of Gothenburg	University of Valencia	
University of Groningen	University of Valladolid	
University of Ioannina	University of Warmia and Mazury in Olsztyn	
University of Kassel	University of Warsaw	
University of Konstanz	University of Western Brittany	
University of L'Aquila	University of Wroclaw	
University of Limerick	University of Zagreb	
University of Lleida	University of Zaragoza	
University of Luxembourg	Vilnius University	
University of Maribor	Warsaw University of Life Sciences	
University of Messina	Warsaw University of Technology	
University of Modena and Reggio Emilia	West Pomeranian University of Technology, Szczecin	
University of Montpellier	Wroclaw Medical University	
University of Muenster	Wroclaw University of Science and Technology	
University of Naples Federico II	Wroclaw University of Environmental and Life Sciences	
University of Padova		
University of Pavia		
University of Perugia		
University of Poitiers		
University of Rennes 1		
University of Strasbourg		
University of Stuttgart		
University of the Balearic Islands		
University of Trento		
University of Twente		
University of Veterinary Medicine Vienna		
University of Vienna		

Uppsala University		
Utrecht University		
Vita-Salute San Raffaele University		
Vrije Universiteit Amsterdam		
Wageningen University and Research		

SM 3 List of Italian universities according to the two main factors of the Principal component analysis

University	factor 1	factor 2
Brescia University	-0,50797	-1,00319
Gabriele D'Annunzio University	0,08411	-0,44614
Politecnico di Bari	-0,73952	2,83376
Politecnico di Milano	-0,09548	2,11853
Politecnico di Torino	-0,59691	0,16824
Roma Tre University	-0,30095	0,70842
Sapienza University of Rome	-1,12927	-0,13678
Università Cattolica del Sacro Cuore	0,81817	-0,50095
Università del Salento	-0,22745	-0,39205
Università Politecnica delle Marche	1,29874	0,02424
University of Bari Aldo Moro	-0,55753	0,90573
University of Bologna	0,04147	-1,35906
University of Cagliari	-1,21717	-1,83525
University of Calabria	-0,13615	0,06285
University of Campania “Luigi Vanvitelli”	0,61295	0,69513
University of Catania	2,36044	0,11739
University of Ferrara	-0,63029	-1,48866
University of Florence	0,90894	0,09144
University of Genoa	-0,37447	-1,29873
University of Insubria	0,55927	-0,43568
University of L'Aquila	-0,73436	0,62997
University of Messina	0,15682	0,13403
University of Milan	-0,59423	-1,12944
University of Milano-Bicocca	0,71889	0,06891
University of Modena and Reggio Emilia	1,26267	-1,44072
University of Naples Federico II	1,15882	2,44879
University of Padova	1,17686	0,20757
University of Palermo	-2,14184	1,1005
University of Parma	-0,4144	-0,1081
University of Pavia	1,2336	-0,05775
University of Perugia	0,66259	-0,32965
University of Pisa	0,19814	-1,2066

University of Rome Tor Vergata	-0,84616	-0,21849
University of Salerno	-0,1804	0,91316
University of Sassari	-1,87882	0,56994
University of Siena	-0,50425	-0,43571
University of Trento	1,21476	0,98792
University of Trieste	-1,37245	-0,5121
University of Turin	-0,16788	0,19325
University of Udine	-1,67165	0,04319
University of Verona	1,02585	-0,04977
Vita-Salute San Raffaele University	1,52652	-0,63813

SM 4: Politecnico di Bari Information retrieval from WOS

To analyze the strategies adopted in the different FOS and the topics addressed by scholars of the Politecnico di Bari, we downloaded the complete list of all publications (including articles and review articles) in English attributed to Politecnico di Bari on the Web of science, with publication date in the years from 2016 to 2019.

The query performed on 18/11/2021 at 15.30 on WOS core collection is as follows:

Politecnico di Bari *(Affiliation) and* ***2016*** *or* ***2017*** *or* ***2018*** *or* ***2019*** *(Publication Years) and* ***Articles*** *or* ***Review Articles*** *(Document Types) and* ***English*** *(Languages)*

From WOS, with the previous query, we identified 2296 publications. After downloading all the papers information from Web of Science, this information and the FOS assigned to each publication by LR were merged via UT code using the mapping file provided by LR on its official webpage (https://www.leidenranking.com/information/fields). Following the methodology described on the LR website for the assignment of the single article to the various FOS to which it belongs, we calculated the number of publications normalized by FOS of Politecnico di Bari. The results obtained are presented in Table 4A and Table 4B. Minor discrepancies are present between the data we retrieved according to the data processing described above and what is reported in the official LR s, mostly due to the WoS database update.

Table 4A. Allocation of FOS of the publications of Politecnico di Bari

FOS	FOS	FOS	Number of publications	Fractional count of publications per FOS
BIO	LIFE		1	0.5
BIO	MAT		11	0.5
BIO	FIS		12	0.5
BIO	SSH		1	0.5
BIO			47	1
LIFE	BIO		3	0.5
LIFE	MAT	FIS	1	0.33
LIFE	MAT		8	0.5
LIFE	FIS		43	0.5
LIFE	SSH		8	0.5
LIFE			101	1
MAT	BIO		5	0.5
MAT	LIFE		5	0.5
MAT	FIS	LIFE	1	0.33
MAT	FIS		95	0.5
MAT	SSH		31	0.5
MAT			188	1
FIS	BIO	LIFE	1	0.33
FIS	BIO		19	0.5
FIS	LIFE		56	0.5
FIS	MAT		116	0.5
FIS	SSH		1	0.5
FIS			1251	1
SSH	LIFE	FIS	1	0.33
SSH	MAT	BIO	1	0.33
SSH	MAT		27	0.5
SSH	FIS		2	0.5
SSH			73	1

Table 4B. Total publications (fractional count) by FOS of the Politecnico di Bari

	Total publications (fractional count) by FOS
BIO	73.66
LIFE	164.33
MAT	338
FIS	1424.33
SSH	108.66