

Statistical evaluation of research performance of young university scholars: A case study

Avaliação estatística do desempenho em pesquisa de jovens estudantes universitários: estudo de caso

Petr PRAUS¹  0000-0002-1336-4108

Abstract

The research performance of a small group of 49 young scholars, such as doctoral students, postdoctoral and junior researchers, working in different technical and scientific fields, was evaluated based on 11 types of research outputs. The scholars worked at a technical university in the fields of Civil Engineering, Ecology, Economics, Informatics, Materials Engineering, Mechanical Engineering, and Safety Engineering. Principal Component Analysis was used to statistically analyze the research outputs and its results were compared with factor and cluster analysis. The metrics of research productivity describing the types of research outputs included the number of papers, books and chapters published in books, the number of patents, utility models and function samples, and the number of research projects conducted. The metrics of citation impact included the number of citations and h-index. From these metrics – the variables – the principal component analysis extracted 4 main principal components. The 1st principal component characterized the cited publications in high-impact journals indexed by the Web of Science. The 2nd principal component represented the outputs of applied research and the 3rd and 4th principal components represented other kinds of publications. The results of the principal component analysis were compared with the hierarchical clustering using Ward's method. The scatter plots of the principal component analysis and the Mahalanobis distances were calculated from the 4 main principal component scores, which allowed us to statistically evaluate the research performance of individual scholars. Using variance analysis, no influence of the field of research on the overall research performance was found. Unlike the statistical analysis of individual research metrics, the approach based on the principal component analysis can provide a complex view of the research systems.

Keywords: Metrics research. Multivariate analysis. Research performance evaluation. Young scholars.

Resumo

O desempenho da pesquisa de um pequeno grupo de 49 jovens, estudantes de doutoramento, pesquisadores júniores e de pós-doutorado, que atuam em diferentes campos técnicos e científicos, foi avaliado com base em 11 tipos de resultados de pesquisa. Os estudantes desenvolvem o seu trabalho de pesquisa numa universidade técnica nos campos de Engenharia Civil, Ecologia, Economia, Informática, Engenharia de Materiais, Engenharia Mecânica e Engenharia de Segurança. Uma avaliação estatística dos resultados da pesquisa foi realizada por análise de componentes principais, e seus resultados foram comparados com a análise de fatores e agrupamentos. As métricas da produtividade da pesquisa que descrevem os tipos de resultados de pesquisa incluíram o número de artigos, livros e capítulos de livro publicados, o número de patentes, modelos de utilidade e amostras de função e o número de projetos de pesquisa conduzidos. As métricas de impacto da citação incluíram o número de citações e o índice-h. A partir dessas métricas –

¹ VŠB – Technical University of Ostrava, Faculty of Metallurgy and Materials Engineering, Department of Chemistry, Listopadu 15, 70830, Ostrava-Poruba, Czech Republic. E-mail: <petr.praus@vsb.cz>.

Received on January 27, 2017, final version resubmitted on July 11, 2017 and approved on August 16, 2017.

Como citar este artigo/How to cite this article

Praus, P. Statistical evaluation of research performance of young university scholars: A case study. *Transformação*, v. 30, n. 2, p. 167-177, 2018. <http://dx.doi.org/10.1590/2318-08892018000200003>



variáveis –, a análise de componentes principais extraiu quatro principais componentes. O primeiro componente principal caracterizou as publicações citadas em jornais indexados pela Web of Science. O segundo componente principal representou os resultados da pesquisa aplicada, e o terceiro e o quarto componentes principais representaram outros tipos de publicações. Os resultados das análises de componentes principais foram comparados com o método hierárquico de análise de clusters de Ward. Os gráficos de dispersão de análises de componentes principais e as distâncias de Mahalanobis, calculadas a partir das quatro pontuações de Componentes Principais, permitiram avaliar estatisticamente o desempenho da pesquisa individual dos estudantes. Usando a análise de variância, nenhuma influência do campo de pesquisa no desempenho geral da pesquisa foi encontrada. Ao contrário da análise estatística das métricas de pesquisa individuais, esta abordagem baseada em análise de componentes principais pode fornecer uma visão complexa dos sistemas de pesquisa.

Palavras-chave: Métricas de pesquisa. Análise multivariada. Avaliação do desempenho da pesquisa. Jovens estudantes.

Introduction

The performance of university scholars is under evaluation on several occasions, such as during the selection of candidates for teaching and research positions, selection of research projects for funding, evaluation of effectiveness of study and research programs, etc. These decisions should be made objectively and transparently based on various reliable and objective criteria.

Research performance can be evaluated by the metrics of productivity and citation impact (Mingers; Leydesdorff, 2015). The first category consists of different metrics, such as the number of papers, books, reports, chapters in books, the number of patents and utility models, the number of granted research projects, the amount of funds obtained for research, etc. The second category of metrics includes the number of citations, the h-index (Hirsch, 2005) and other related indexes (Mingers; Leydesdorff, 2015). Possibilities of measuring research productivity is a broad topic, which has been widely discussed in many papers for a long time, e.g. (Nagpaul; Roy, 2003; Abramo; Cicero; D'Angelo, 2013; Abramo; D'Angelo, 2014).

At technical universities, technical and scientific fields based on applied and fundamental research and their outputs can hardly be compared with each other. For example, the results of engineering disciplines are mostly patented contrary to those in economics or computer science, which are mostly published in papers. The number of papers and citations indicate the scientific impact of researchers in their field (Podlubny, 2005; Podlubny; Kassayova, 2006). The h-index is used to measure productivity, as well as the number of citations of researchers, and it can be also used to evaluate departments, universities, and research institutes (Lazaridis, 2010). Although the h-index is often criticized, particularly due to the dependence on age of the researchers and type of scientific fields, it is still evaluated by different databases, such as the Web of Science (WoS), Scopus and Google Scholar. Some modifications in the h-index have been developed (Schreiber; Malesios; Psarakis, 2012).

The results of applied research are mostly characterized by the number of patents, utility models, software, prototypes, function samples, etc. These results are mostly developed in cooperation with industrial partners by contractual research and have a direct impact on the development of society. The number of granted projects is not a typical research performance metrics, but it shows the ability of scholars/researchers to set up and conduct projects on new and attractive research topics. These projects are often granted to persons and teams who have been producing high-quality results that are appreciated by the research community. In general, the research performance is a complex multivariate problem, which must be solved by appropriate statistical methods.

The Principal Component Analysis (PCA) is a basic multivariate statistical method that is used to reduce dimensionality of original data described by several variables to few main significant components (Jolliffe, 2002). PCA is used to solve many scientific and technical problems that are described by many variables. In the case of research evaluation, PCA has already been applied to analyze university rankings (Dehon; McCathine; Verardi, 2010; Cova; Pais; Formosinho, 2013; Docampo; Cram, 2015), and evaluate pharmaceutical firms (Ramani, 2002), research institutes (Ortega; López-Romero; Feranández, 2011), different research teams (Rey-Rocha; Garzón-García;

Martín-Sempere, 2006), or other scientific fields (Almeida; Pais; Formosinho, 2009). PCA was also used to describe country participation in 6th Framework Programme (Ortega; Aguillo, 2010). In the literature, there are some PCA applications to discriminate different kinds of scientists characterized by several metrics including the h-index and g-index (Costas; Bordons, 2008) to analyze 39 measures of scientific impact (Bollen *et al.*, 2009), bibliometrics, and scientific metrics (Franceschet, 2009; Costas; Van Leeuwen; Bordons, 2010; Todeschini, 2011; Scheriber; Malesios; Psarakis, 2012; Torres-Salinas *et al.*, 2013), as well as the patenting activities of scholars (Baldini; Grimaldi; Sobrero, 2007) etc.

As given above, PCA has been widely used in Scientometrics, but there are only few applications to evaluate small groups of researchers. The aim of this study was to evaluate the research performance of a small group of 49 young scholars, whose outputs were produced in different technical and scientific fields and in different variability resulting from different work experiences. The researchers consisted of doctoral students, postdoctoral, and junior researchers working at a technical university. The performance of each scholar was represented by 11 metrics consisting of several outputs of fundamental and applied research and citation metrics. To evaluate the overall research performance, the Mahalanobis distances were used to identify each scholar and were calculated from selected principal component scores and statistically processed. This approach is new and it allows us to compare scholars working in different fields.

Methodological procedures

Data collection

The data analyzed was composed of the research metrics of 49 young scholars in fundamental and applied research carried out at a technical university in the Czech Republic. Their basic research outputs were represented by the number of papers published in peer-reviewed journals with impact factors indexed in WoS (Jimp) and without impact factors (Jrev) indexed in other databases, the number of papers published in Conference Proceedings (CP), the number of Books (B) and Chapters published in books (CH), and the number of Research Projects (RP) obtained. The quality of their research work was expressed by the h-index (HI) and the total number of Citations (Cit). The conference proceedings were indexed by WoS. The total number of citations and h-index were also evaluated using WoS. The applied research outputs were characterized by the number of Patents (PT), Utility Models (UM), and Function Samples (FS).

The basic statistics of the research metrics were expressed by their total number (N), mean, standard deviation, and maximal magnitudes (Table 1). The original data matrix contained 51 scholars, but 2 of them were considered as outliers by the box-and-whisker diagrams and excluded from further statistical treatment. These 2 scholars had an excessive number of citations (752) and utility models (18), respectively. The 49 scholars analyzed worked in 7 different research fields: Civil Engineering ($n=4$), Ecology ($n=16$), Economics ($n=7$), Computer Science ($n=7$), Materials Engineering ($n=10$), Mechanical Engineering ($n=3$), and Safety Engineering ($n=2$).

Theory

Principal Component Analysis

The main objective of PCA is to look for new latent (hidden) variables of n samples, which are orthogonal (not correlated) to each other. Each latent variable – principal component – is a linear combination of p variables x and it describes a different source of total variation:

$$t1m = x_{i1} w_{1m} + x_{i2} w_{2m} + \dots + x_{ip} w_{pm} \quad (1)$$

Where t_{im} is the score of the i -th object in the m -th component. The component loadings are the contribution measures of a particular variable to the principal components. The variability of the principal components is given by corresponding eigenvalues λ_m , where $m=1,2, \dots, p$, which are ordered as $\lambda_1 > \lambda_2 > \dots > \lambda_p$, where each eigenvalue is the variance of the corresponding m -th component. PCA can be performed by the eigenvalue decomposition of a correlation (or covariance) matrix or by the singular value decomposition of the original data matrix.

Cluster analysis

Cluster analysis consists of a number of different methods that organize objects into groups of similar objects. This exploratory method is used to discover the data structure among not only observations, but also among variables, arranged into dendrograms. The utilized methods, algorithms, and similarity/dissimilarity measures are described elsewhere in the literature (Everitt, 2001). In this study, common linkage (single-linkage, complete-linkage, average linkage) and the Ward's hierarchical clustering (Ward, 1963) method were used for the analysis of research performance.

Statistic computations

The original data matrix was set up and processed in MS Excel. Multivariate analysis and ANOVA and other statistical calculations were performed by the software packages STATGRAPHIC Plus 5.0, QC.Expert 3.3.0.6. (Trilobyte) and XLSTAT 2017 (Addinsoft). Before the multivariate analysis, the data were standardized to avoid misclassifications arising from different orders of magnitude of the variables. For this purpose, the data were mean centered (μ) and scaled by the standard deviations (σ) as $(x-\mu)/\sigma$.

Results and Discussion

The research performance of the young scholars was characterized by the 11 metrics shown in Table 1. The metrics were selected to cover various disciplines, in which the fundamental and applied research was performed. The metrics showing publication activities of the scholars were the number of papers published in journals and conference proceedings, and the number of books and chapters published in books. The applied research results were characterized by the number of patents, utility models, and function samples. The quality of research work was expressed by the h-index, the number of citations of the overall scientific production, and the number of research projects conducted. The articles in journals with impact factor, citations, and h-index were taken from WoS, which has been preferred by the Czech research evaluation system (Vaněček, 2014; Good *et al.*, 2015). The peer-reviewed journals were taken from other databases, such as Scopus, ERIH, and a list of the Czech non-impact and peer-reviewed journals (R&D..., 2015).

Principal component analysis was performed by computing the eigenvalues and corresponding eigenvectors of the correlation matrix composed of the abovementioned research metrics. Prior to PCA, the Bartlett's Sphericity test confirmed that the correlation matrix was significantly different from the identity one ($p < 0.0001$). There is no universal rule for estimating the number of PCs. The eigenvalues of all PCs were calculated at 2.9866, 1.9727, 1.5726, 0.9778, 0.8939, 0.7556, 0.6814, 0.62269, 0.2982, 0.1612 and 0.0773. The corresponding cumulative variabilities were calculated at 27.15%, 45.09%, 59.38%, 68.27%, 76.40%, 83.27%, 89.46%, 95.12%, 97.83%, 9.30% and 100.00%, respectively. According to the magnitude of eigenvalue, which should be equal to or higher than 1 (Kaiser, 1960), 3 main PCs explaining 59.00% of the total data variance can be selected. However, this total variance is low, so a scree plot was used to estimate the number of appropriate PCs (Cattell, 1966). The steep eigenvalue decrease stopped at the 4th PC, therefore, 4 main PCs providing 68.00% of the total data variance were selected: 1st PC (PC1), 2nd (PC2), 3rd

PC (PC3) and 4th PC (PC4) explained 27.20%, 17.90%, 14.30% and 8.90% of the total variance, respectively. Thus, the original data dimensionality was reduced from 11 to 4. Relationships among the research metrics were discussed and the individual scholars were visualized by means of the PC scatter plots.

Interpretation of principal components

In general, the interpretation of principal components is necessary to understand the data structure. The component loadings (Table 2) were considered in order for us to find relationships among original metrics.

The first PC was mostly influenced by the number of articles in journals with impact factors, number of their citations, and the authors' h-index. It is not surprising that all these metrics correlated well with each other because they usually indicate high-quality publication results. The highest variability of these metrics was given by the high differences among individual scholars publishing in journals with impact factors.

The second PC was mostly saturated by the outputs of applied research (function samples, utility models, and patents) and the number of book chapters. Unlike Jimp, the number of all these outputs were relatively low (Table 1) and were produced by several authors. Therefore, the variation of PC2 was lower. The function samples were the prevailing applied research output. Their number well correlated with the number of utility models because they were mostly created by the same scholars. The number of patents and utility models also significantly

Table 1. Basic statistics of scholars' outputs.

	Jimp	Jrev	CP	B	CH	RP	HI	Cit	PT	UM	FS
N	363	653	424	87	61	188	--	1303	10	10	32
Mean	7.41	13.3	8.65	1.78	1.24	3.84	2.41	26.6	0.20	0.20	0.65
SD	6.96	12.9	18.1	2.60	1.90	3.86	1.81	41.3	0.61	0.58	1.41
Max	33	60	121	13	10	15	9	233	3	3	6

Note: CP: Conference Proceedings; B: Number of Books; CH: Chapter published in books; RP: Research Projects; HI: h-index; Cit: Citations; PT: Patents; UM: Utility Models; FS: Function Samples.

Source: Prepared by author (2017).

Table 2. Loadings of 4 main principal components.

Metrics	PC1	PC2	PC3	PC4
B	0.2556	-0.2672	0.2919	-0.1052
Cit	0.5538	-0.0259	-0.0398	0.0211
CP	0.0116	0.2188	-0.4774	-0.5561
FS	0.0434	0.4200	0.1784	0.0480
HI	0.5322	0.0541	-0.1499	0.4991
CH	-0.0628	-0.4229	0.0732	0.0380
Jimp	0.5156	0.0545	-0.1910	-0.2211
Jrev	-0.0054	-0.2616	0.3321	0.4965
PT	-0.0083	0.3540	0.4283	0.2677
RP	0.2607	-0.2188	0.4184	0.0322
UPT	0.0702	0.5267	0.3470	0.2416

Note: PC: Principal Component.

Source: Prepared by author (2017).

correlated with each other for the same reason. PC2 also contained a relatively high loading of the book chapters, which negatively correlated with the applied research results. The authors who published the chapters were not those who were producing patents and other applied results.

The third PC was mainly influenced by the number of papers in peer-reviewed journals without impact factors, the number of papers in conference proceedings, books and research projects. The low variability of the metrics, such as *Jrev* and *CP*, indicated that a majority of the scholars published in these media. Their negative correlation indicated that the authors were deciding between both possibilities. A few books were published by several authors mostly in Czech publishing houses. Therefore, this metric showed low variation and was included in PC3. The number of research projects was correlated with the number of books because both outputs mostly had the same authors. The high positive loading of the patents (0.4283) cannot be explained by their relationships with these metrics. A probable explanation is that the research metrics of low magnitudes and low variations weakly correlate with each other without any logical reason.

The fourth PC, just as PC3, was saturated by the number of papers in peer-reviewed journals and number of papers in conference proceedings, but the *h-index* also influenced this component. However, the correlation coefficients between these variables were too low (up to -0.204) to come to any conclusion.

In general, the main PCs were found to characterize well the articles cited in journals with impact factor (PC1), the results of applied research (PC2), and other types of publications (PC3 and PC4). The PCA results were confirmed by factor analysis, which identified similar factor loadings (not shown here).

Evaluation of scholars in PC space

The scholars were evaluated according to their performance using the PCA scatter plots. The scatter plot composed of the PC1 and PC2 scores is shown in Figure 1a. The high positive PC1 scores indicate many well-cited papers published in high-impact journals, which is typical for scholars working in the fields of Materials Engineering, Ecology, and Economics. The high positive PC2 scores represent the high number of applied research results, but the high negative PC2 scores indicate a lot of chapters in books. Therefore, the points in the 1st quadrant represent the scholars with the best performance. The scholars placed in the 4th quadrant were good at publishing and those in 2nd quadrant created more outputs of applied research outputs. Most of the scholars in the 3rd quadrant created relatively the low numbers of papers as well as applied research outputs. Obviously, the scholars nº 31, 41 and 42 significantly differ from the others due to their high number of articles in high-impact journals and, consequently, the high number of citations and *h-indexes*. The scholar nº 1 is different due to the higher number of patents, utility models, and function samples. The scholars nº 12 and 30 published a lot of book chapters.

Figure 1b shows the scatter plot of the PC1 and PC3 scores. The points for numbers 1, 31, 41 and 42 are also well visible. The scholar nº 3 published many contributions in the conference proceedings. The scholars with high negative PC3 scores published in conference proceedings. Likewise, most of the scholars are concentrated in the 2nd and 3rd quadrants. The two scatter plots also show that no well-defined clusters consisting of the scholars from the same or similar fields were found.

As discussed above, all the scholars can be characterized by the coordinates in the PC space. Their locations in the scatter plot quadrants as well as their distances from the coordinate origins are important for their evaluation. To improve resolution of individual persons, the Mahalanobis distance *T* for each scholar was calculated as follows:

$$T_i = \sqrt{(X_i - \mu)^T C^{-1} (X_i - \mu)} \quad (3)$$

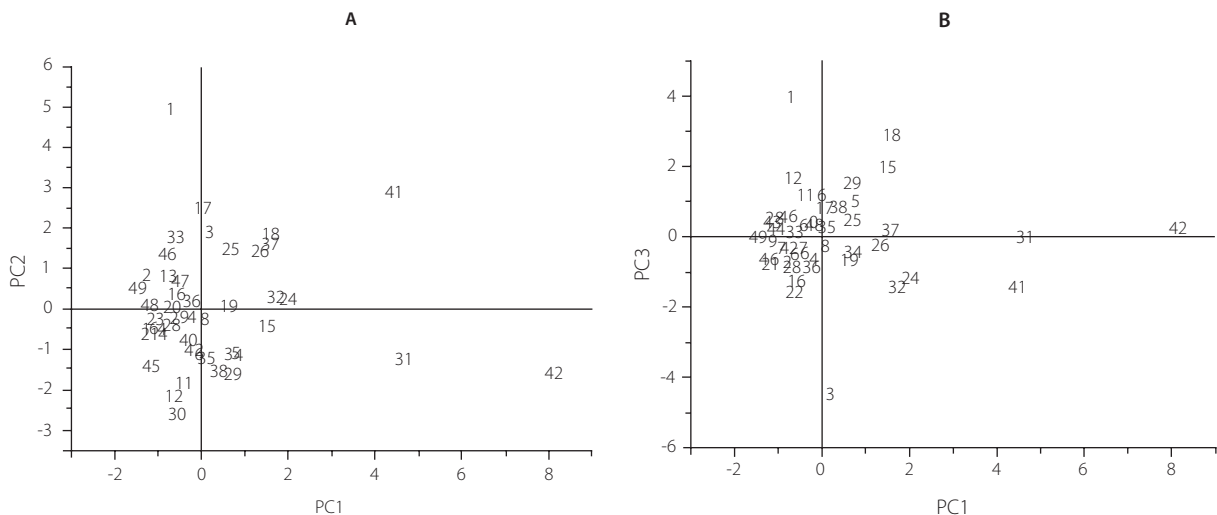


Figure 1. Scatter plot of PC1 and PC2 scores.

Source: Prepared by author (2017).

Where μ is the mean of the PC score x_i , C is the correlation matrix composed of the PC scores (Maesschalck; Jouan-Rimbaud; Massart, 2000), and m is the number of principal components ($m=4$). Statistical tests of the T values of all 49 scholars showed the lognormal distribution. Therefore, logarithms of the distances T were calculated to obtain the normal distribution (Skewness=0.596, Kurtosis=3.85) for further statistical use; the normality was also confirmed by the D'Agostino ($p=0.196$) and Kolmogorov-Smirnov ($p=0.133$) tests. The overall performance chart was developed as shown in Figure 2.

The mean $\mu_T=0.157$ and standard deviation $\sigma_T=0.233$ of $\log T$ were calculated. The performance limit $\mu_T+2\sigma_T$ ($=0.623$) separates the scholar with excellent or weak results. The overall performance of the scholars between μ_T and $\mu_T+\sigma_T$ ($=0.390$) should be better or worse than that of the average scholars who can be found below μ , that is, close to the origin of the PC space. The exceptional scholars, numbers 1, 3, 18, 31, 41, 42, previously identified in the scatter plots are well visible. The overall performance chart can be also used as a control chart for the statistical evaluation of scholars over a long course of time.

The Mahalanobis distances in the research fields were also tested by one-way ANOVA. The variance tests, such as Cochran's C ($p=0.397$), Bartlett's ($p=0.570$), Hartley's, Levene's ($p=0.982$), Kruskal-Wallis ($p=0.514$) and Scheffé's tests ($p=0.941-1.000$) confirmed that there were no statistically significant differences between the standard deviations of the T magnitudes corresponding to the fields. In addition, the multivariate ANOVA (MANOVA) was applied to the original data. Wilks's test (Rao's approximation) confirmed that there was no significant effect of the fields on the research outputs ($p=0.403$).

Comparison of Principal Component Analysis with cluster analysis

The PCA results were compared with hierarchical clustering of the original 11-dimensional data. The linkage clustering methods were tested, but the best organized dendrograms were obtained by Ward's method.

The dendrogram in Figure 3 shows two main clusters corresponding to the publication characteristics and the applied research characteristics. The left "publication" cluster is divided into two sub-clusters: the publications in impact journal and the citation characteristics similar to PC1 and other publications similar to PC3. The composition of the right "applied research cluster" is similar to PC2. Unlike PCA, the book chapters were included in the

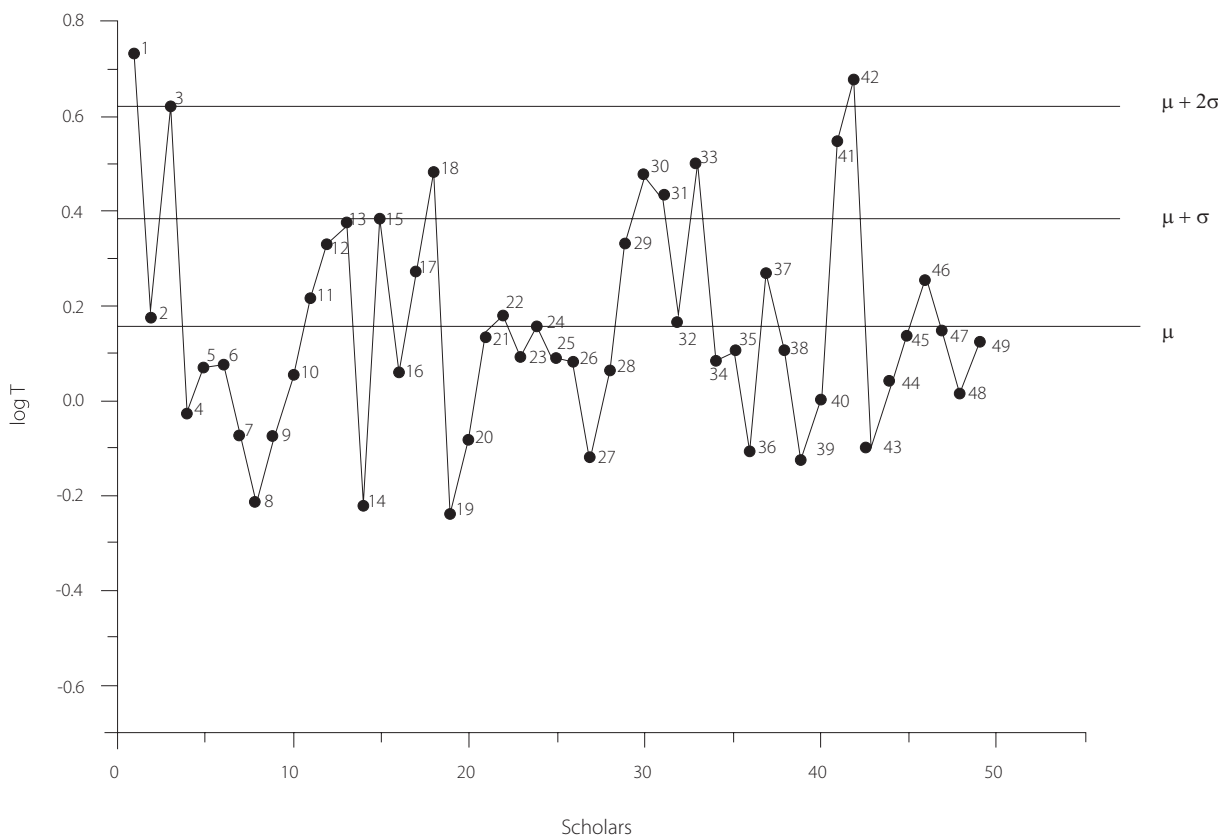


Figure 2. Overall research performance chart of scholars evaluated.

Source: Prepared by author (2017).

“publication” cluster”. The reasons may be due to the different mechanisms of the two methods and the presence of information noise, which was removed by PCA.

Two main clusters are visible in the dendrogram in Figure 3b. The composition of the clusters was compared with the points distributed in quadrants of the scatter plot shown in Figure 1a. The left cluster contains the scholars from the 1st quadrant and a few scholars from the 2nd and 4th quadrants. The right large cluster can be divided into two sub-clusters combining the scholars from the 2nd, 3rd and 4th quadrants. Ward’s method was also applied to the clustering of the scholars exhibited in the PC space, but the composition of clusters was mixed. In addition, the separation of scholars into clusters was independent on their performance.

The dendrogram of the research outputs corresponded well with the results of the PCA. The PCA scatter plots were more transparent and better organized than the dendrograms. The position of scholars in the quadrants can easily explain their activities: the publication of articles in high-impact journals connected with citation impact (PC1), the production of applied research outputs (PC2) and other publications (PC3 and PC4). Better resolution of individual scholars was achieved using their overall performance chart.

Scholar’s evaluation in context with research metrics

The PCA results provided an insight into the structure of the young scholars working in different research fields. It was found that the publication of papers in high-impact and well-cited journals was the main problem

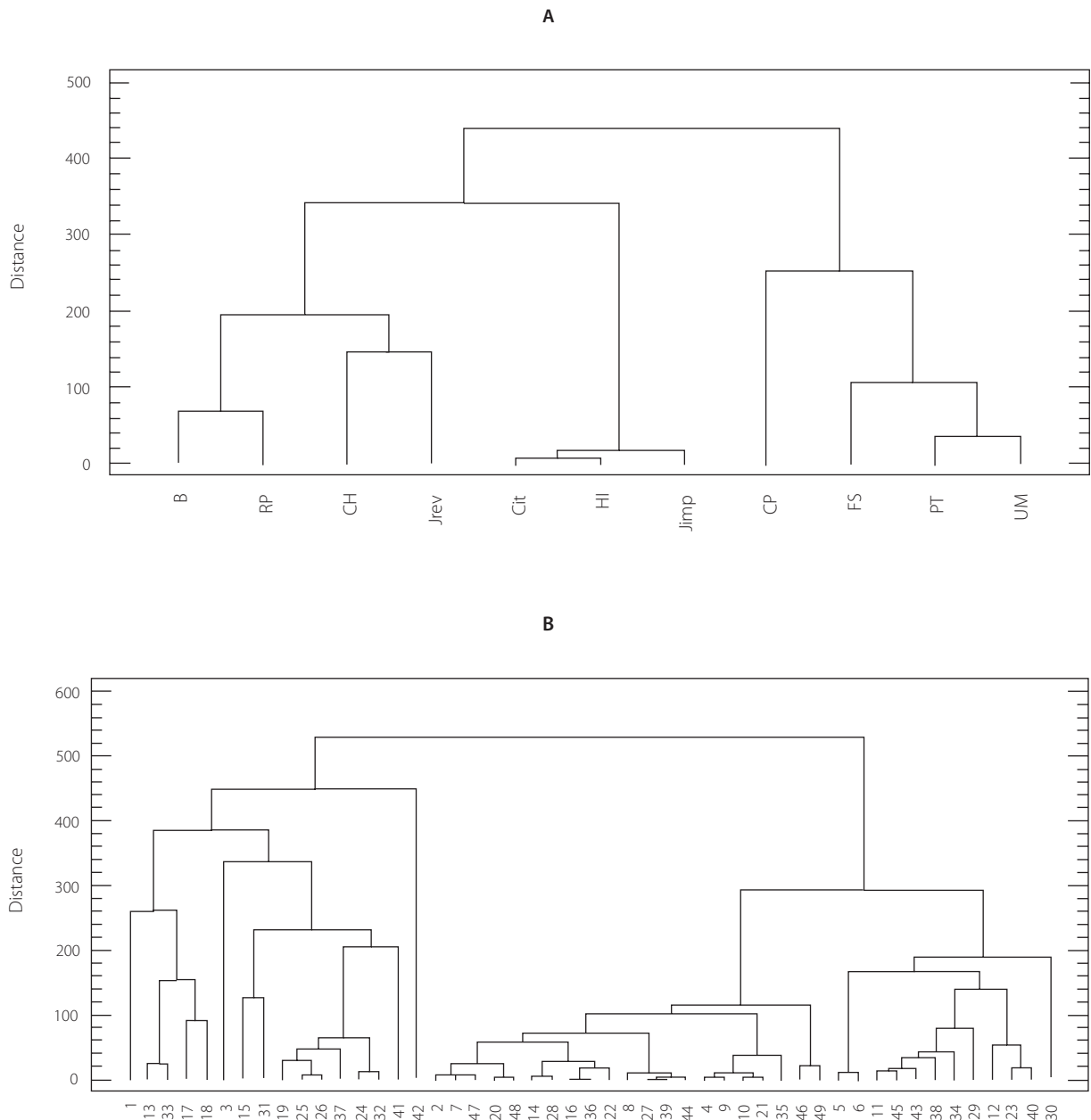


Figure 3. Ward's dendrogram of research outputs and Ward's dendrogram of scholars.

Source: Prepared by author (2017).

indicated by the highest variation. The scholars must be motivated to conduct high-quality research whose results could be published in well-cited scientific journals (Vieira; Gomes, 2010). It is one of the ways to improve the overall research performance and the quality of the educational system.

On the other hand, the articles in conference proceedings were frequent results. There are two main reasons for this. First, in comparison with high-impact scientific journals, the demands on the quality of conference papers are lower. This effect is evident especially in cases of large international multidisciplinary conferences. Second, the publication process of conference proceedings is much faster than in scientific journals. These factors play a role in rapidly developing fields, such as engineering sciences and computer science (Larsen; Ins, 2010).

The lower numbers of patents, utility models, and function samples indicate that these results are difficult to obtain because cooperation with industrial partners is needed. In addition, the actual system of research evaluation and funding in the Czech Republic evaluates only patents (breeds and varieties) and underestimates projects of contractual research (Good *et al.*, 2015). Therefore, scholars and researchers prefer the production of patents to other applied research outputs.

The problems mentioned above cannot be solved without the effective motivation of scholars and the intensification of domestic and international cooperation with other universities, research institutions, and industrial partners. The development of networks with new partners could reveal new opportunities for research cooperation and stimulation of research performance.

Conclusion

The research performance of 49 young scholars was characterized by 11 metrics of productivity and citation impact. The productivity metrics were selected to describe multidisciplinary fundamental and applied research conducted in different research fields in departments of a technical university. The metrics of impact were the number of citations and the h-index. The metrics of productivity involved the number of papers in journals with and without impact factors, the number of papers in conference proceedings, the number of books and chapters in books, the number of patents, utility models and function samples, and the number of research projects conducted. In general, there are other outputs of applied research, such as the number of prototypes, applied and certified methods, software and industrial designs, but they were not used as metrics because only few scholars developed them.

Principal component analysis reduced these 11 original metrics to 4 main PCs explaining 68% of the total data variance. PC1 characterized the cited publications in impact journals indexed by WoS. PC2 represented the results of applied research, while PC3 and PC4 represented other types of publications. PC1 indicated that the main problem of the overall research performance was the low productivity of high-quality papers that could be published in well-cited and high-impact journals.

The PCA scatter plots enabled the visualization of scholars according to their research performance. In addition, the application of the Mahalanobis distances calculated from the scores of the 4 main PCs enabled the statistical evaluation of overall research performance of individual scholars. Using ANOVA of the Mahalanobis distances it was found that the scholars' results did not depend on their specialization. The PCA results were compared with the hierarchical clustering (Ward's method).

The principal component analysis enabled the evaluation of a small group of young scholars with relatively low number of outputs in different technical and scientific fields and indicated the weak and strong points of a university research system. This PCA-based approach can provide a complex view on the research performance characterized by various metrics. It can be applied for easy and reliable evaluation of multidisciplinary research carried out by research groups and individuals at universities and other research institutes.

References

Abramo, G.; Cicero T.; D'Angelo C. A. Individual research performance: A proposal for comparing apples to oranges. *Journal of Infometrics*, v. 7, n. 2, p. 528-539, 2013.

Abramo, G.; D'Angelo C. A. How do you define and measure research productivity? *Scientometrics*, v. 101, n. 2, p. 1129-1144, 2014.

Almeida, J. A. S.; Pais, A. A. C. C.; Formosinho, S. J. Science metrics and science patterns in Europe. *Journal of Infometrics*, v. 3, n. 2, p. 134-142, 2009.

Baldini, N.; Grimaldi, R.; Sobrero, M. To patent or not to patent? A survey of Italian inventors on motivations, incentives, and obstacles to university patenting. *Scientometrics*, v. 70, n. 2, p. 333-354, 2007.

- Bollen, J. *et al.* A principal component analysis of 39 scientific impact measures. *PLoS One*, v. 4, n. 6, p. e6022, 2009. Available from: <<http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006022>>. Cited: May 3, 2018.
- Cattell, R. B. The scree test for the number of factors. *Multivariate Behavioral Research*, v. 1, n. 2, p. 245-276, 1966.
- Costas, R.; Bordons, M. Is g-index better than h-index? An exploratory study at the individual level. *Scientometrics*, v. 72, n. 2, p. 267-288, 2008.
- Costas, R.; Van Leeuwen T. N.; Bordons, M. A bibliometric classificatory approach for the study and assessment of research performance at the individual level. *Journal of the American Society for Information Science and Technology*, v. 61, n. 8, p. 1564-1581, 2010.
- Cova, T. G. G. G.; Pais, A. A. C. C.; Formosinho, S. J. Iberian universities: A characterisation from ESI rankings. *Scientometrics*, v. 94, n. 3, p. 1239-1251, 2013.
- Dehon, C.; McCathine, A.; Verardi, V. Uncovering excellence in academic rankings: A closer look at the Shanghai ranking. *Scientometrics*, v. 83, n. 2, p. 515-524, 2010.
- Docampo, D.; Cram, L. On the effects of institutional size in university classification: The case of the Shanghai ranking. *Scientometrics*, v. 102, n. 2, p. 1325-1346, 2015.
- Everitt, B. Cluster Analysis. 4th ed. London: Hodder Arnold, 2001.
- Franceschet, M. A Cluster analysis of Scholar and Journal Bibliometric Metrics. *Journal of the American Society for Information and Technology*, v. 60, n. 10, p. 1950-1964, 2009.
- Good, B. *et al.* Counting quality? The Czech performance-based research funding system. *Research Evaluation*, v. 24, n. 2, p. 91-105, 2015.
- Hirsch J. E. An index to quantify an individual's scientific research output. *Proceeding of the National Academy of Sciences*, v. 102, n. 46, p. 16569-16572, 2005.
- Jolliffe, I. T. Principal component analysis. 2nd ed. New York: Springer-Verlag, 2002.
- Kaiser, H. F. The application of electronic computers to factor analysis. *Educational and Psychological Measurement*, v. 20, n. 1, p. 141-151, 1960.
- Larsen, P. O.; Ins, M. The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, v. 84, n. 3, p. 575-603, 2010.
- Lazaridis, T. Ranking university department using mean h-index. *Scientometrics*, v. 82, n. 2, p. 211-216, 2010.
- Maesschalck, R.; Jouan-Rimbaud, D.; Massart, D. L. The mahalanobis distance. *Chemometrics and Intelligent Laboratory Systems*, v. 50, n. 1, p. 1-18, 2000.
- Mingers, J.; Leydesdorff, L. A review of theory and practice in scientometrics. *European Journal of Operational Research*, v. 246, n. 1, p. 1-19, 2015.
- Nagpaul P. S.; Roy S. Constructing a multi-objective measure of research performance. *Scientometrics*, v. 56, n. 3, p. 383-402, 2003.
- Ortega, J. L.; Aguillo, I. F. Describing national science and technology systems through a multivariate approach: Country participation in the 6th Framework Programme. *Scientometrics*, v. 84, n. 2, p. 321-330, 2010.
- Ortega, J. L.; López-Romero, E.; Feranández, I. Multivariate approach to classify research institutes according to their outputs: The case of the CSIC's institutes. *Journal of Infometrics*, v. 5, n. 3, p. 323-332, 2011.
- Podlubny, I. Comparison of scientific impact expressed by the number of citations in different fields of science. *Scientometrics*, v. 64, n. 1, p. 95-99, 2005.
- Podlubny, I.; Kassayova, K. Towards a better list of citation superstars: Compiling a multidisciplinary list of highly cited scholars. *Research Evaluation*, v. 15, n. 3, p. 154-162, 2006.
- Ramani, S. V. Who is interested in biotech? R&D strategies, knowledge base and market sales of Indian biopharmaceutical firms. *Research Policy*, v. 31, n. 3, p. 381-398, 2002.
- Rey-Rocha, J.; Garzón-García, B.; Martín-Sempere, M. J. Scientists' performance and consolidation of research teams in biology and biomedicine at the Spanish council for scientific research. *Scientometrics*, v. 69, n. 2, p. 183-212, 2006.
- R&D council definition of types of R&D results. [S.l.], 2015. Available from: <<http://www.vyzkum.cz/FrontClanek.aspx?idsekce=733439>>. Cited: June 30, 2017.
- Schreiber, M.; Malesios, C. C.; Psarakis, S. Exploratory factor analysis for the Hirsch index, 17 h-type variants, and some traditional bibliometric metrics. *Journal of Infometrics*, v. 6, n. 3, p. 347-358, 2012.
- Todeschini, R. The j-index: A new bibliometric index and multivariate comparisons between other common indices. *Scientometrics*, v. 87, n. 3, p. 621-639, 2011.
- Torres-Salinas, D. *et al.* On the use of biplot analysis for multivariate bibliometric and scientific metrics. *Journal of the Association for Information Science and Technology*, v. 64, n. 7, p. 1468-1479, 2013.
- Vaněček, J. The effect of performance-based research funding on output of R&D results in the Czech Republic. *Scientometrics*, v. 98, n. 1, p. 657-681, 2014.
- Vieira, E. S.; Gomes, J. A. N. F. Citations to scientific articles: Its distribution and dependence on the article features. *Journal of Infometrics*, v. 4, n. 1, p. 1-13, 2010.
- Ward, J. H. Hierarchical grouping to optimize an objective function. *Journal of the American Statistical Association*, v. 58, n. 301, p. 236-244, 1963.