

## ***Text Mining-based Scientometric Analysis in Educational Research***

Gyula Nagy, University of Szeged, Hungary

The European Conference on Education 2018  
Official Conference Proceedings

### **Abstract**

This paper presents a complete scientometric analysis of Hungarian Pedagogy, the most significant and oldest Hungarian educational research journal, founded in 1892 and still being issued today. In our research project, all of the articles which had been published in this journal (N=6574) was digitised in order to build a well-structured database, which makes it possible to analyse them by means of various metadata. Besides analysing their metadata, our aim was to investigate the full text corpus with text mining, an essential tool of Educational Data Mining. General scientometric indicators and tendencies, such as the amount and length of the articles, the most significant authors' impacts and backgrounds, as well as the number of citations by authors were also discussed. Moreover, we investigated the ratio of male and female authors, the national and institutional background of certain researchers with a full range of metadata analyses. Recent studies have verified that scientific cooperation is growing world-wide; therefore, our first research question focuses on this matter by revealing the co-authorship network of the journal. The hubs of this graph represent the most central people in the collaborative authorship of the analysed journal in the field of Hungarian educational research. Finally, after a co-authorship graph had been created, enormous citation graphs were also produced based on the analysed journal in order to reveal the scientific network within the field of educational research in Hungary. Using this graph, a multi-criteria citation analysis was conducted enabling the indication of additional relevant results.

Keywords: text mining, scientometrics, educational research journals, co-authorship network, citation analysis, citation network

**iafor**

The International Academic Forum  
[www.iafor.org](http://www.iafor.org)

## Introduction

Scientometric databases (also known as citation databases) were introduced several years ago (Garfield, 2006). In the beginning, only the so-called 'hard' sciences took advantage of the benefits thereof, but today social sciences and humanities also build upon the opportunities offered by citation databases to a certain extent (Hammarfelt, 2016). Despite these possibilities, many scientific publications exist whose influence and their authors' scientific impact cannot be measured due to the fact that an abundance of journals are not represented in international citation databases. These academic journals mainly derive from the area of humanities and social sciences (Mongeon & Paul-Hus, 2016), accordingly, educational research is particularly concerned. These journals – either issued in English or in a non-English language – have two common features: they are all emerging periodicals, and have an important role for a national or a local scientific community.

The following two questions truly highlight the seriousness of the problem: How many scientific publications are born globally?; How many of them are contained in Web of Science (hereinafter referred to as WoS) or Scopus databases? Naturally, we cannot give an exact answer for the first question, but the order of magnitude is certainly much larger than the numbers in the second answer. To get closer to the problem, international academic journal indexing databases (such as Ulrich, DOAJ, ERIH Plus, etc.) could be good bases for comparison.

The national current research information systems could be other viable benchmarks for comparison. Sivertsen (2014) suggested really poor coverage of some fields of science regarding the two leading scientometric providers according to the data of the Current Research Information System in Norway (Cristin). Their results show that Scopus covers 32%, while WoS covers only 23% of all peer reviewed scholarly articles published by Norway's higher education institutions in journals and series of the field of humanities. The comparable figures for the social sciences are 54% in Scopus and 42% in WoS. The corresponding figures in the examined citation databases for the education and educational research category are 35% and 22%. Unfortunately, in many countries the coverage situation is similar to or worse than in the Norwegian example.

All this proves that there is a huge gap in science evaluation, therefore we should find alternative solutions. Although our results are related to one well-selected journal only, both the problem and the applied method are truly universal, as we can see from the examples which will be detailed later on. Actually, we suggest that the major outcome of our research is not the presented results themselves but rather the developed and applied method which could be adapted successfully by anyone in similar instances. Taking all these into consideration, the main question is: How can those journals be measured which are out of the scope of the major international scientometric providers? Somehow, we have to reproduce the technique of the huge content providers in a smaller dimension. This paper reports on such experiment in the area of educational research, with scientometrics, citation and text mining analysis.

## **Theoretical background**

The abundance of digitized contents which became available on the Internet due to the transition from an analogue to a digital environment has brought about a remarkable revolution in recent decades in the scientific communication (Moss & Endicott-Popovsky, 2015). In recent years, many authors have dealt with this transformation (Castelli, Manghi, & Thanos, 2013; Pearce, Weller, Scanlon, & Kinsley, 2010). In addition to many other tendencies, one characteristic of this metamorphosis is that scientific communication has increasingly shifted from the world of books (monographs, edited and synthesized volumes) to journal papers and various conference publications. This change has taken place even in social sciences or arts and humanities (Bornmann & Mutz, 2015; Larsen & von Ins, 2010), similarly to the previously experienced in 'hard' sciences.

It is obvious that we are close to making everything available online, either in a digitised or born-digital form (Nagy, 2014). Along with commercial databases, institutional repositories have a crucial role in this process (MIT, 2016). Repositories also support the Open Access movement, serving its green route (Björk, Laakso, Welling, & Paetau, 2014). The emergence of institutional and disciplinary repositories and the transformation of the publishing model of scientific journals were absolutely fundamental in the change taking place in scientific communication in the last decades (Chan, 2004; Lynch & Lippincott, 2005). Evidently, these effects have also concerned educational researchers. As it was mentioned earlier, recent studies suggest that scientific collaboration is growing world-wide (Adams, 2012), and the significance of scientific evaluation has been considerably increasing in all fields of science. New viewpoints and methods have emerged (Hicks, Wouters, Waltman, De Rijcke & Rafols, 2015). Nevertheless, such adaptation is required in all disciplines, in educational research, too.

The study which deals with measuring, analysing and evaluating science is called scientometrics. Its subjects can cover a broad spectrum; ranging from an individual to a group of scholars, institutions, journals, and so on. In fact, it is an outstandingly innovative discipline with a lot of new approaches (e.g. altmetrics). It is important to note that in scientometrics various fields of science are handled in different ways: 'hard' sciences, social sciences, arts and humanities have very different points of view (Nederhof, 2006; Archambault, Vignola-Gagné, Côté, Larivière, & Gingrasb, 2006).

After all, more and more disciplines or interdisciplines try to exploit full-text corpora, and it is necessary to take advantage of the same within the discipline of scientometrics as well. This procedure is fundamentally based on the solutions offered by artificial intelligence, using the tools of data and text mining. Moreover, these methods and other IT-provided solutions are soon going to be built into the toolkit of every discipline. Text mining provides a unique opportunity for investigating latent information in texts (Zhang, Porter, & Chiavetta, 2017), especially in scientific papers (Liu, Yu, Janssens, Glänzel, Moreau, & De Moor, 2010). Text mining is the process of deriving information from the text, therefore it would be obviously useful in an educational context, as well, where lots of texts are produced day by day. For instance, it could be applicable for educational log file analysis, finding citations in educational research journals, content analysis, and so on.

As it was previously pointed out, lots of disciplines use data and text mining; moreover, each has created its own applied version. Educational Data Mining (EDM) is one of the best examples of the manifestation of this phenomenon. EDM is an emerging discipline used for exploring large-scale data and textual information coming from some educational settings (Baker & Yacef, 2009). This interdisciplinary method has been increasingly applied in educational research, and several publications have been already published by the experts of EDM with the use of some kind of text mining tools (e.g. Koedinger, D'Mello, McLaughlin, Pardos, & Rosé, 2015; Wang, Bowers, & Fikis, 2017). The International Educational Data Mining Society was founded in 2011 and the first conference of the EDM experts dates back to 2008, from then on the conference has been organized every year.

### **Aims and research questions**

Our research project is a pilot study aimed at experimenting and introducing new methods to scientometrics for those journals which are not indexed by internationally acknowledged citation databases. The subject of our research is a nationally acclaimed educational research journal entitled “Hungarian Pedagogy”. This scientific journal is over a hundred years old (it was founded in 1892), and it has still been a leading journal in the area of Hungarian educational research. First of all, we had to digitize all the volumes of the journal; then we had to build a bibliographic database from the metadata for the subsequent research.

Beyond that, our main aim was to carry out a highly detailed scientometric analysis with creating the co-authorship graph and generating the inner and full citation network of the analysed journal. These co-authorship graph and citation networks reveal the scientific network within the field of educational research in Hungary for the first time, as similar graphs have not been produced beforehand. A more general goal was to develop new tools that could serve the interest of academics: to investigate their scientific performance mainly in the area of educational research and their scientific communication in the field of education. The following research questions were raised in this paper:

- (1) How have the main scientometric indicators (number of publications; average length of studies; average number and freshness of references per article) changed over time and how have the patterns of co-authorship transformed?
- (2) Who are the most influential authors of the journal, and what is their institutional background like?
- (3) Has the ratio of male and female authors changed in the last 25 years?

### **Methods**

First, in an attempt to create an operable and universal method, we had to select an experimental subject from the area of educational research. Our choice was “Hungarian Pedagogy” (in Hungarian: Magyar Pedagógia), the most important and oldest Hungarian educational research journal, representing a wide range of acknowledged authors. After all of its volumes (114) had been digitized, a 50,000-page text corpus was created from the journal articles (ca. 6,500 scientific studies). In addition, all metadata was put into a well-structured database, which having been built by cataloguing librarians.

The full-text corpus was analysed with text mining, a method which can show latent information and connections in the text structure. Along with the text mining analyses, a scientometric analysis was also conducted. The research is a pilot study aimed at the scientometric and text-mining analyses of national journals from the area of social sciences and humanities. These two disciplines were chosen due to the fact that presently they are much less represented in the well-known international citation databases (WoS, Scopus).

To carry out the citation analyses, it was necessary to retrieve and structure the references from the articles. It could be only managed with an automatized approach, due to the abundant number of elements. To implement this step, we had to convert the texts from PDF to XML format. A crucial element in the project was the automatic detection and extraction of citations, which became possible with the applied standard reference style (it is very similar to the well-known APA style). The pattern recognition was conducted by using text mining, with exploiting the help of regular expressions: we have taken into consideration the special additional style information, such as bold and cursive characters, which added more data for us.

All of the metadata and the extracted citations have been stored in structured CSV files. To handle these raw data, we used LibreOffice for manipulation and Microsoft Office for visualization. Furthermore, to analyse and visualise the co-authorship and citation graphs in an accurate way, we applied Gephi, an open-source software package for network analysis and visualization. In the course of our work, we had to tackle lots of problems and challenges; for instance, the non-consistent compliance of the official citation style, misspellings and errors or missing data elements. In spite of all these difficulties, according to our estimation at least 95% of all references could be extracted. After this step, citations were verified and manual data cleaning was conducted. Altogether, exactly 14,039 citations were identified for subsequent examinations.

## **Results and Discussion**

### **General scientometric profile**

Firstly, in 2015 the full text of all the volumes was digitized and made freely available for the scientific community in an institutional repository. The digitized texts contribute to the content of the repository, which consists of thousands of articles. These articles are accessible for the experts of the particular area making research on the history of education easier (with the assessment of old issues), increasing their visibility and enabling the dissemination of current issues. In addition, it also opens the opportunity for computerized analysis of the texts. In spite of its high academic quality, Hungarian Pedagogy is not represented in international citation databases; therefore, our work served as a niche for the national scientific evaluation in educational research.

Secondly, a full range of metadata analyses were performed by assessing top authors, group of authors, co-authors, female authors, the average length of articles and many other aspects (Nagy & Molnár, 2017). In addition, a co-authorship and a reference graph (from 1991 to 2014) were built up, and a multi-criteria citation analysis was

conducted. Results indicate the most-cited authors, interdisciplinarity of the articles, international and national references and their freshness. We give a general overview of the scientometric profile of the analysed journal. Based on our results, the main tendencies are the following: the number of publications has been decreasing but the average length of the articles has been continuously increasing. The ratio of female authors has enhanced, similarly to the proportion of articles with multiple authorship. Finally, according to our investigation results, the average number of references per article has also ascended.

The findings of the present paper do support the claims according to which the productivity of the Hungarian educational researchers has been decreasing in the past decades. If we compare our results to other topic-specific national journals and international databases, we can see a general reduction in the number of publications. Parallely, the number of the Hungarian educational researchers' international publications has only slightly increased in the last decades. At the same time, according to further outcomes, a new phenomenon occurred concerning the female authors' predominance, as the ratio of male and female authors has turned around since the millennium (see Figure 1), answering our third research question. This investigation was also carried out with taking into account the entire publishing period of the journal. It is scarcely surprising that these results show greater changes since the end of the 19<sup>th</sup> century.

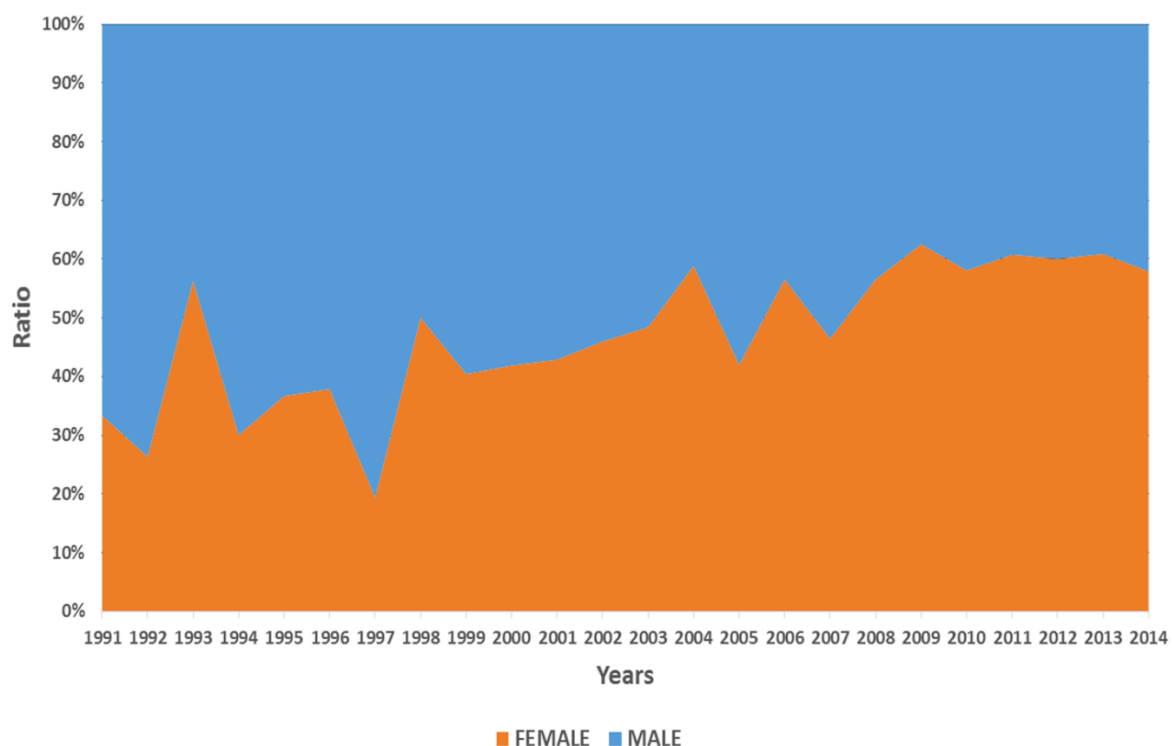


Figure 1: Rise in the number of female authors in the last 25 years in the analysed educational research journal

According to our results based on the average length of the publications, the amount of references, their freshness and the intensifying collaboration between the authors via co-authorship patterns, the scientific quality of the examined journal has

significantly increased in the last twenty-five years. The average number of references per publications has changed from around 20 to around 60 over the last quarter of a century. Besides, in the second half of the examined period, authors have undisputedly cited more recent articles than in the nineties. Since 2002, the average freshness of citations per year is between 11 and 18 years, while this value was about 10 years more in the previous decade. Based on these results, it can be concluded that the quantity of publications decreased but their quality significantly increased during the investigated period.

In the next step, besides assigning the general indicators, we had to categorize and typify citations for further specific examinations. This task was not a trivial operation due to the large number of elements and the various literature sources which were used by the authors. To identify the titles of Hungarian journals, we used the Hungarian Periodicals Table of Contents Database (in Hungarian: MATARKA), while with regard to the international journals, we used the Journal Citation Reports. Afterwards, we supplemented it by filtering typical terms that are specific to certain journal titles (e.g. quarterly, bulletin, etc.). The classification of the journals was done by assigning them to the following main categories: periodical or non-periodical and Hungarian or international. So, we created groups with various settings, such as Hungarian and international journals. Then, the formerly formed groups could be subdivided into subgroups, such as journals with impact factor or the subgroup of educational research journals. With this, the purpose was to examine the proportions of each category in order to draw different conclusions from them and establish predominant trends in the distribution of the referred sources. The most important results can be found in Table 1.

Table 1: The 30 most popular sources of literature in Hungarian Pedagogy between 1991 and 2014

#	<i>Title</i>	<i>Number of citations</i>	<i>Language</i>	<i>Genre</i>
1.	Magyar Pedagógia [Hungarian Pedagogy]	434	Hungarian	journal
2.	Iskolakultúra [School Culture]	398	Hungarian	journal
3.	Új Pedagógiai Szemle [New Pedagogical Review]	271	Hungarian	journal
4.	Journal of Educational Psychology	114	English	journal
5.	Educatio	106	Hungarian	journal
6.	Pedagógiai Szemle [Pedagogical Review]	97	Hungarian	journal
7.	Az iskolai tudás [School Knowledge]	94	Hungarian	book
8.	Köznevelés [Public Education]	72	Hungarian	journal
9.	Learning and Instruction	56	English	journal
10.	Journal of Personality and Social Psychology	53	English	journal
11.	Az iskolai műveltség [Literacy at School]	52	Hungarian	book
12.	Magyar Pszichológiai Szemle [Hungarian Psychological Review]	52	Hungarian	journal
13.	Az olvasási képesség fejlődése és fejlesztése [The development of reading ability]	49	Hungarian	book
14.	Psychological Review	46	English	journal
15.	Pszichológia [Psychology]	44	Hungarian	journal

16.	Child Development	41	English	journal
17.	Contemporary Educational Psychology	41	English	journal
18.	Handbook of Self-Regulation	41	English	book
19.	Jelentés a magyar közoktatásról [Report on Hungarian Public Education]	38	Hungarian	book
20.	Review of Educational Research	37	English	journal
21.	A biológia tanítása [Biology Teaching]	36	Hungarian	journal
22.	American Psychologist	35	English	journal
23.	Educational Psychologist	35	English	journal
24.	Science Education	33	English	journal
25.	Modern Nyelvoktatás [Modern Language Education]	33	Hungarian	journal
26.	Tanulmányok a neveléstudomány köréből [Studies in the Field of Pedagogy]	32	Hungarian	book
27.	XXI. század és nevelés [Education and the 21 <sup>st</sup> century]	32	Hungarian	book
28.	Neveléstudomány az ezredfordulón [Educational Sciences at the Turn of the Millenium]	32	Hungarian	book
29.	Modern Language Journal	31	English	journal
30.	Developmental Psychology	30	English	journal

A more detailed examination was required in those cases when authors have cited journal articles, consequently, a sub-sample had to be composed from the full sample. To this end, the proportion of the usage of international and Hungarian literature sources was also reported. In this design, a 43%-57% distribution could be detected in the Hungarian-international dimension. According to this data, we can claim that international references are predominant. Taking into account the temporal distribution of the data, we can state that the references of international journals have an increasing emphasis in the used resources, as from 2007 onwards the annual distribution of international journals is constantly between 60% and 70%. The growing proportion of international journal references also predicts that authors are increasingly citing more publications with impact factor in their studies. The upward trend shown in Figure 2 clearly confirms this premise.



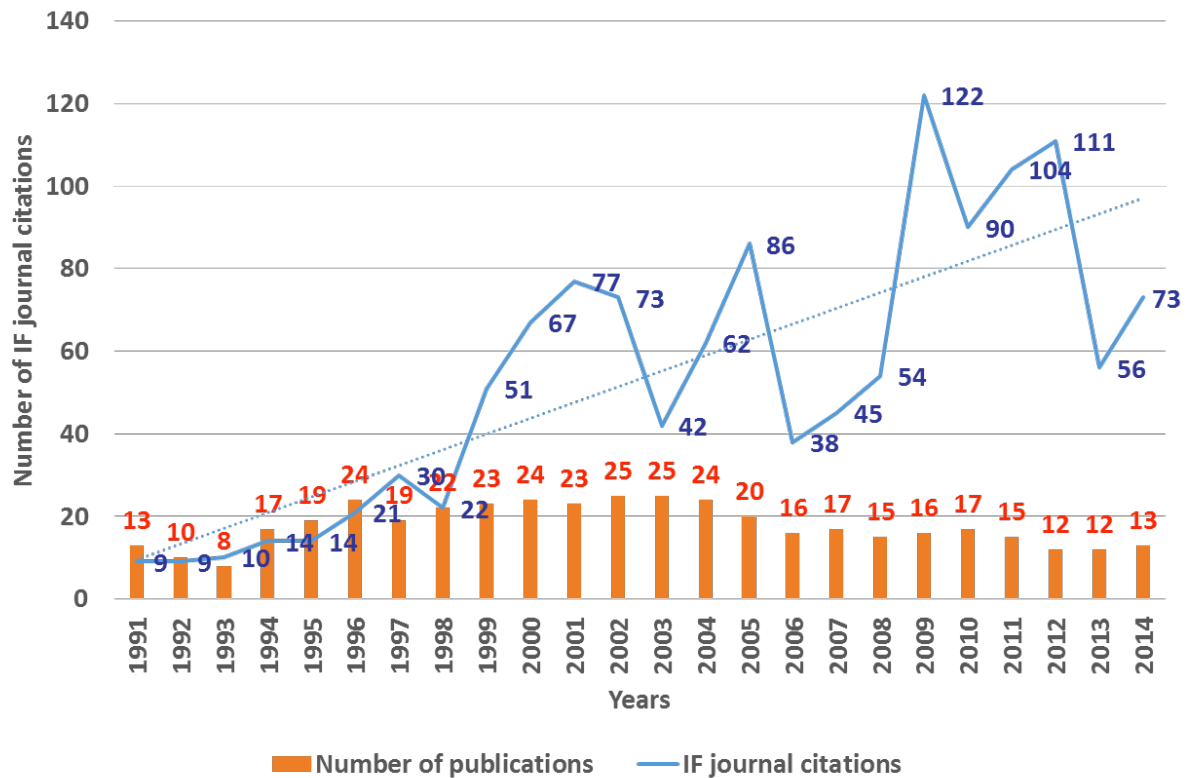


Figure 2: References to journals with impact factor in the light of the annual publication number

### Co-authorship and citation network analysis

As far as the collaboration between academics is concerned, the results justify the developing scientific cooperation among educational researchers due to the evolving co-authorship of Hungarian Pedagogy. On average, more and more authors have written an article as time went by during the last quarter of a century. In the beginning of the examined period, the average author per article values were closer to one author per article, i.e. the predominant proportion of the articles was written by one author, but in the second period these values were closer to 1.5. In one year it reached a value of 1.77, which is outstanding compared to all of the other examined years.

Figure 3 illustrates a typical co-authorship pattern from the area of Hungarian educational research, represented by a part of the created co-authorship graph. The circles symbolize authors, with their size marking their importance in the collaboration network, while their colour indicating their institutional background, partly responding to our second research question. The whole co-authorship graph consists of 108 larger or smaller distinct components. This number implies a strong fragmentation for a graph of such size.

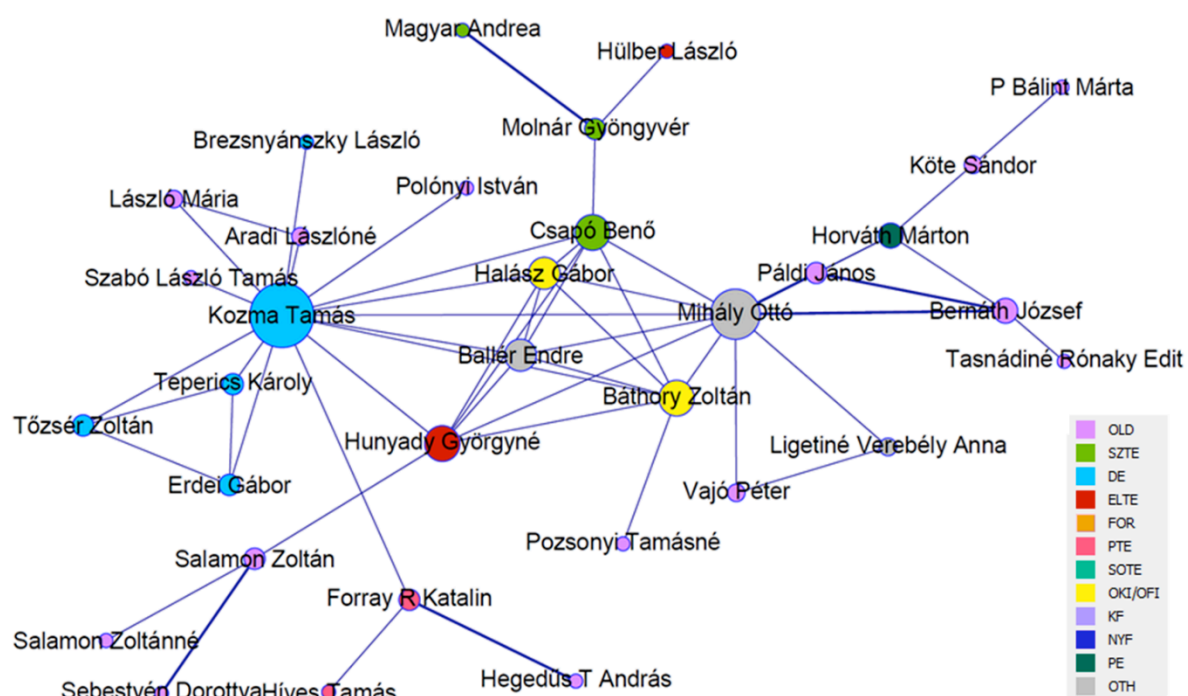


Figure 3: The largest component from the co-authorship graph

After creating the collaboration network, citation networks were also implemented. The plural form is actually not unintentional because we have created two types of citation graphs. The first one is an inner network only with Hungarian Pedagogy citations. This graph consists of 222 nodes, 467 edges and has 15 separated components. Beside the inner citation network, the full citation network was also completed and resulted in the following, most important general metrics: the number of nodes is 10,382, while the number of edges is 19,182. The average degree of the network is 3.695 and the network diameter is 9. The number of connected components is 6; however, a large component typifies the citation graph with 10,245 nodes. Last but not least, the visualization of the full citation network shows a really nice graphical illustration from a distant view. Because there are a large number of elements, it can only be interpretable in detail in a special software environment.

In addition to the previous outcomes, the scrutinized journal's most cited authors and their institutional affiliation were identified (see Figure 4), responding again to our second research question. In most cases, the obtained results in connection with the author's impact are consistent with the indicators of the international reference databases. The internationally recognized scholars in the area of Hungarian educational research are the most cited researchers in the articles of Hungarian Pedagogy as well. Moreover, the results in connection with the institutional affiliation show some surprising facts; for instance, the educational researchers of only two major Hungarian universities, the University of Szeged (22.07%) and Eötvös Loránd University (16.22%) are mainly represented as the most-cited authors.

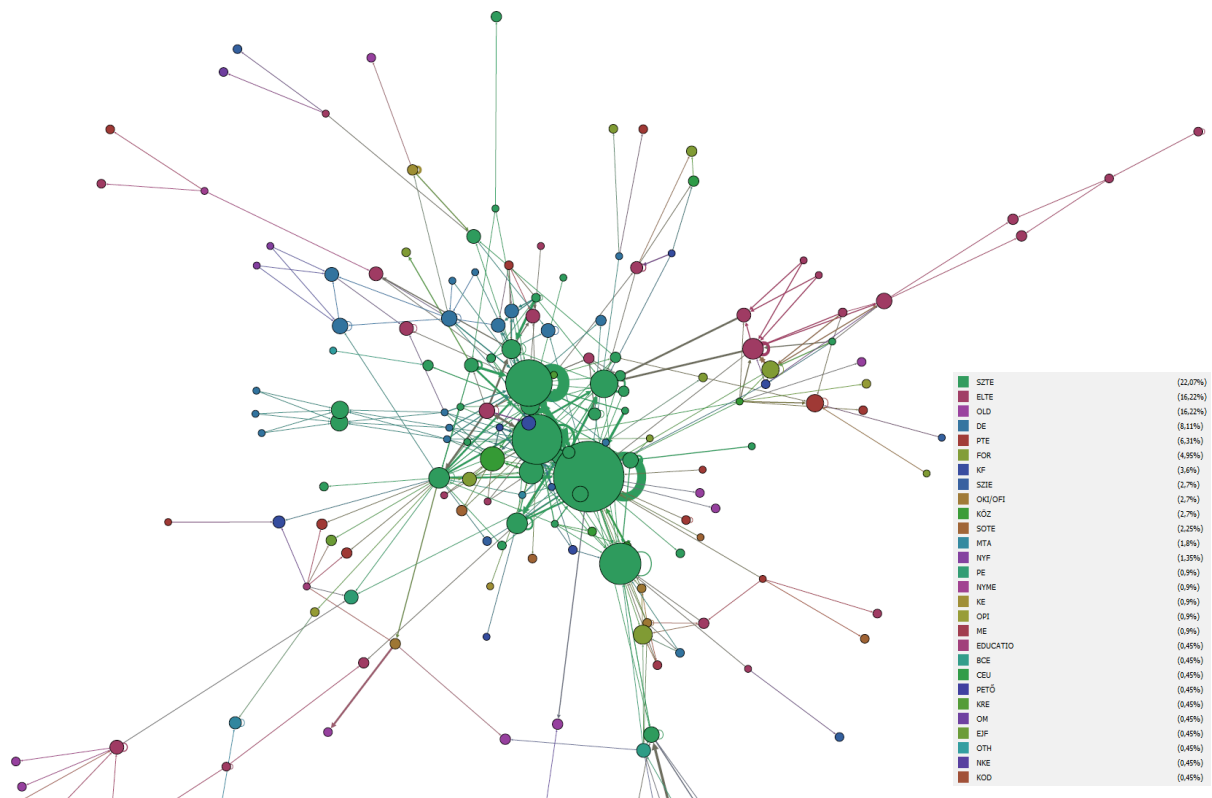


Figure 4: The internal citation graph with the correspondent institutional affiliation

## Conclusions

This pilot study was conducted for making up a shortage of in relation to a special claim in the area of science evaluation in such cases when the international scientometric databases cannot be used. It would like to provide a practical and unique response to a common question, as the measurement of the scientific quality of many scientific media proves to be particularly difficult. This problem can originate from a multitude of reasons; for instance, they might be newly founded journals only with a few years history or they may be issued in a native language with specific but relevant topics for a local scientific community. Nevertheless, these journals have a really important role at national level consequently they deserve special attention with regard to their scientometric analysis.

Our study reveals a scientometric profile of an acknowledged Hungarian educational research journal, the Hungarian Pedagogy. The obtained results completely corroborated our presumptions in the investigated period (1991-2014). The number of publications has been decreasing but the average length of the articles has been continuously increasing. The average number of references per article has also ascended and the authors have cited more up-to-date publications in the second half of the examined period. A very similar quality change has started in 2007, i.e. nearly two-thirds of the journal citations involved international articles since then. Finally, in the last twenty-five years the ratio of female authors has enhanced, similarly to the proportion of articles with multiple authorship.

It is more important, however, that a detailed scientometric profile has been completed of an essential educational research journal at national level. As our

applied method is widely flexible and adaptable, it may serve as a schema for further comparative analyses concentrating on similar research questions. The limitations of the technique might derive from the required IT expertise which may not always be readily available. The non-consistent compliance of the official citation style, misspellings and errors or missing data elements in the reference lists may cause further difficulties even in the process of the automatic extraction of citations. Altogether, the conducted research demonstrates that scientometrics is relevant and provides beneficial possibilities in educational research. This means that there is life beyond citation databases: alternative solutions enable us to evaluate those journals which are not indexed by these databases.

Besides the relevance of scientometrics, another important outcome is that text mining has a great potential in the area of educational research, so it is advised for scholars to take advantage of this opportunity. In connection with scientific publications, lots of hidden information can be extracted from the descriptive metadata by using bibliographic databases. For example, a full co-authorship network of a journal can be created only from the usage of metadata. Building collaboration and citation networks also have great potential, and there are some really useful open source software packages for this purpose.

### **Acknowledgements**

This study was conducted with the financial support from OTKA K115497.

## References

- Adams, J. (2012). Collaborations: The rise of research networks. *Nature*, 490(7420), 335-336. <https://doi.org/10.1038/490335a>
- Archambault, É., Vignola-Gagné, É., Côté, G., Larivière, V., & Gingras, Y. (2006). Benchmarking scientific output in the social sciences and humanities: The limits of existing databases. *Scientometrics*, 68(3), 329-342. <https://doi.org/10.1007/s11192-006-0115-z>
- Baker, R. S., & Yacef, K. (2009). The state of educational data mining in 2009: A review and future visions. *JEDM| Journal of Educational Data Mining*, 1(1), 3-16.
- Björk, B.-C., Laakso, M., Welling, P., & Paetau, P. (2014). Anatomy of green open access. *Journal of the Association for Information Science and Technology*, 65(2), 237-250. <https://doi.org/10.1002/asi.22963>
- Bornmann, L., & Mutz, R. (2015). Growth rates of modern science: A bibliometric analysis based on the number of publications and cited references. *Journal of the Association for Information Science and Technology*, 66(11), 2215-2222. <https://doi.org/10.1002/asi.23329>
- Castelli, D., Manghi, P., & Thanos, C. (2013). A vision towards Scientific Communication Infrastructures: on bridging the realms of Research Digital Libraries and scientific Data Centers. *International Journal on Digital Libraries*, 13(3-4), 155-169. <https://doi.org/10.1007/s00799-013-0106-7>
- Chan, L. (2004). Supporting and enhancing scholarship in the digital age: the role of open access institutional repository. *Canadian Journal of Communication*, 29(3), 277-300. <https://doi.org/10.22230/cjc.2004v29n3a1455>
- Garfield, E. (2006). The history and meaning of the journal impact factor. *JAMA*, 295(1), 90-93. <https://doi.org/10.1001/jama.295.1.90>
- Hammarfelt, B. (2016). Beyond coverage: Toward a bibliometrics for the humanities. *Research Assessment in the Humanities* (pp. 115-131). Springer, Cham. [https://doi.org/10.1007/978-3-319-29016-4\\_10](https://doi.org/10.1007/978-3-319-29016-4_10)
- Hicks, D., Wouters, P., Waltman, L., De Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, 520(7548), 429-431. <https://doi.org/10.1038/520429a>
- Koedinger, K. R., D'Mello, S., McLaughlin, E. A., Pardos, Z. A., & Rosé, C. P. (2015). Data mining and education. *Wiley Interdisciplinary Reviews: Cognitive Science*, 6(4), 333-353. <https://doi.org/10.1002/wcs.1350>
- Larsen, P. O., & von Ins, M. (2010). The rate of growth in scientific publication and the decline in coverage provided by Science Citation Index. *Scientometrics*, 84(3), 575-603. <https://doi.org/10.1007/s11192-010-0202-z>

Liu, X., Yu, S., Janssens, F., Glänzel, W., Moreau, Y., & De Moor, B. (2010). Weighted hybrid clustering by combining text mining and bibliometrics on a large-scale journal database. *Journal of the Association for Information Science and Technology*, 61(6), 1105-1119. <https://doi.org/10.1002/asi.21312>

Lynch, C. A., & Lippincott, J. K. (2005). Institutional repository deployment in the United States as of early 2005. *D-Lib Magazine*, 11(9), 1-11.

Massachusetts Institute of Technology (2016). *Institute-wide Task Force on the Future of Libraries: Preliminary Report*. Retrieved from <https://future-of-libraries.mit.edu/sites/default/files/FutureLibraries-PrelimReport-Final.pdf> - Available: 9 July 2018

Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: a comparative analysis. *Scientometrics*, 106(1), 213–228. <https://doi.org/10.1007/s11192-015-1765-5>

Moss, M., & Endicott-Popovsky, B. (Eds.). (2015). *Is Digital Different? How information creation, capture, preservation and discovery are being transformed*. Facet Publishing.

Nagy, G. (2014). Megy-e a digitalizálás által a világ előbb? Avagy mi végre digitalizálunk?. *Információs Társadalom*, 14(3), 44-52.

Nagy, G., & Molnár, G. (2017). A Magyar Pedagógia folyóirat tudománymetriai elemzése: tendenciák, szerzők, társszerzőségi együttműködés. *Magyar Pedagógia*, 117(1), 5-27. <https://doi.org/10.17670/mped.2017.1.5>

Nederhof, A. J. (2006). Bibliometric monitoring of research performance in the social sciences and the humanities: A review. *Scientometrics*, 66(1), 81-100. <https://doi.org/10.1007/s11192-006-0007-2>

Pearce, N., Weller, M., Scanlon, E., & Kinsley, S. (2010). Digital scholarship considered: How new technologies could transform academic work. *In education*, 16(1), 33–44.

Sivertsen, G. (2014). Scholarly publication patterns in the social sciences and humanities and their coverage in Scopus and Web of Science. *Proceedings of the science and technology indicators conference* (pp. 598-604).

Wang, Y., Bowers, A. J., & Fikis, D. J. (2017). Automated Text Data Mining Analysis of Five Decades of Educational Leadership Research Literature: Probabilistic Topic Modeling of EAQ Articles From 1965 to 2014. *Educational Administration Quarterly*, 53(2), 289-323. <https://doi.org/10.1177/0013161X16660585>

Zhang, Y., Porter, A. L., & Chiavetta, D. (2017). Scientometrics for tech mining: an introduction. *Scientometrics*, 111(3), 1875-1878. <https://doi.org/10.1007/s11192-017-2344-8>

**Contact email:** [gyula.nagy@ek.szte.hu](mailto:gyula.nagy@ek.szte.hu)