Reference Publication Year Spectroscopy (RPYS) of computer science papers from Eastern Europe

Dalibor Fiala

University of West Bohemia, Plzeň, Czech Republic. Email: dalfia@kiv.zcu.cz.

Lutz Bornmann

Max Planck Society, Munich, Germany. Email: bornmann@gv.mpg.de.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Abstract

Purpose: The current article presents the results of a case study dealing with the historical roots of Eastern European researchers in computer science.

Design/methodology/approach: The study is based on an analysis of cited references stemming from a collection of around 80,000 computer science papers by Eastern European researchers published from 1989 to 2014. By using a method called "reference publication year spectroscopy" (RPYS) for historical analyses based on bibliometric data, we analyze around 800,000 references cited in those papers. We identify the peak years including most frequently cited publications (from 1952, 1965, and 1975) and focus on these outstanding works for the field. We show how these influential papers were cited in Eastern Europe and in general and on which scientific fields they have the most impact.

Findings: A noteworthy publication that seems to have a tremendous effect on Eastern European computer science is Zadeh's "Fuzzy sets" article which appeared in *Information and Control* in 1965. We demonstrate that computer scientists from Eastern Europe are more conservative in their citation behavior and tend to refer to older and more established research than their counterparts from the West.

Originality/value: Which are the historical roots of researchers working in a particular field or on a specific topic? Are there certain publications – landmark papers – which are important for their research? We guess that these are questions bothering researchers in many fields.

Keywords

Reference publication year spectroscopy (RPYS), Citation analysis, Computer science, Eastern Europe, Web of Science

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

1. Introduction

Eastern European countries were isolated to some extent from the Western world for several decades after World War II until 1989 when the so-called Eastern Bloc suddenly collapsed. With the notable exception of the former Yugoslavia, all of these countries were either part of the Soviet Union or were communist regimes under a heavy Soviet influence separated from the West by the Iron Curtain. Against the backdrop of this specific situation, it appears interesting to investigate how current research in Eastern Europe is influenced by the past. Is the past dominated by literature from the Soviet Union or from Western countries? We wonder how computer science research, conducted primarily in the West by scholars in the United States, the United Kingdom, Japan, Canada, and (West) Germany, affected computer scientists in Eastern Europe after the fall of the Berlin Wall. In the following, we take a closer look at the references cited in the papers by Eastern European computer scientists. Using the method Reference Publication Year Spectroscopy (RPYS), we investigate citation patterns, detect the most frequently cited papers, and explore the reasons for their citations. It is specific for RPYS that it is based on cited references data to reveal influential publications.

There has been surprisingly little scientometric research dealing with Eastern Europe in general and with Eastern European computer science in particular. To the best of our knowledge Fiala and Willett (2015) have carried out the most comprehensive bibliometric analysis of recent Eastern European computer science so far. Jurajda et al. (2017) have compared the publication performance of selected Eastern and Western European countries. They concluded that the post-communist nations still lag behind. This has been recently confirmed by Prathap (2018) who investigated the research performance of Eastern European universities to show their inferiority to those in Western Europe. Lovakov and Agadullina (2017) have examined the scientific production and collaboration of former Soviet republics in psychology journals. They found that 15 post-Soviet countries produced less than one

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

percent of the world psychology output in the 25 years after the collapse of the Soviet Union. The general decline in scholarly activity in the same geographic area has been explored by Karamourzov (2012). A drastic decrease in research production and innovation results in the whole region was observed in the period 1990-2009.

Grančay et al. (2017) have studied the development of research publication in the field of business and economics in seven Eastern European countries between 2000 and 2015. A four-fold increase in the number of Web of Science (WoS) indexed publications by Central and Eastern European researchers was detected in their study. Kozak et al. (2015) have analyzed research production and international collaboration of former Eastern Bloc countries. They concluded that research performance and international collaboration of these countries improved less than expected after the fall of the communist Bloc. Radošević and Yoruk (2014) have compared the research capacities of nations in Eastern Europe to those of long-time Western European Union member states. They identified Central and Eastern Europe as a "catching-up region" in terms of both research productivity and impact.

In turn, the different approaches to research excellence in the three post-Soviet Baltic countries (Lithuania, Latvia, and Estonia) have been contrasted by Allik (2013). He determined Estonia to be the only one of the three nations that managed to increase its research output as well as impact since gaining independence. Teodorescu and Andrei (2011) have examined the international scientific collaboration of formerly communist European states. They noted that the international collaboration of six Eastern European nations grew substantially after the fall of communism and that internationally co-authored papers (especially with US researchers) had higher citation rates. Apart from analyses of groups of countries, individual nations have been in the focus as well, such as Serbia in Popović et al. (2012) and Ivanović and Ho (2014) or the Czech Republic in Vaněček (2008). None of the papers above, however, has dealt specifically with the references cited in Eastern European

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

computer science publications (and thus the historical roots), which is the concern of the present analysis.

2. Dataset

This study is based on the same dataset as used in Fiala and Willett (2015), which included journal articles, conference papers, and reviews published from 1989 to 2014 and indexed in the WoS database. The indexes searched were the *Science Citation Index Expanded* and the *Conference Proceedings Citation Index – Science*. The research area selected was "Computer Science" and the country field contained 26 post-communist Eastern European states. From the post-Soviet republics we decided to include the three Caucasian nations[1], which lie on the boundary between Eastern Europe and Asia, but not the five Central Asian republics[2], which lie entirely within Asia. We also included the three "container" countries which are now defunct[3]. The final list of inspected countries is thus the following: Albania, Armenia, Azerbaijan, Bosnia and Herzegovina, Bulgaria, Belarus[4], Croatia, Czech Republic, Czechoslovakia, Estonia, Georgia, Hungary, Latvia, Lithuania, Macedonia, Moldova, Montenegro, Poland, Romania, Russia, Serbia, Slovakia, Slovenia, Soviet Union[5], Ukraine, and Yugoslavia.

The records of papers authored by researchers affiliated with the above countries were downloaded from WoS and, after some cleaning, imported in the CRExplorer software (www.crexplorer.net) for further processing (Thor et al., 2016a, 2016b). This software has been especially developed for analyzing historical roots based on cited references data. The dataset then contained 77,395 papers including 824,641 cited references. We restricted the

[1] Armenia, Azerbaijan, and Georgia.
[2] Kazakhstan, Kyrgyzstan, Tajikistan, Turkmenistan, and Uzbekistan.
[3] Czechoslovakia (dissolved on 1 January 1993), Soviet Union (dissolved on 26 December 1991), and Yugoslavia (gradual violent breakup in 1991 and 1992, included only Serbia and Montenegro from 1992 to 2003, officially called Serbia and Montenegro between 2003 and 2006, final dissolution on 5 June 2006).
[4] Called Byelarus in Web of Science.
[5] Called USSR in Web of Science.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

reference publication years to the period between 1900 and 2014. The clustering and merging functionalities of the CRExplorer have been used to clean the cited references dataset by variants (i.e. referenced variants of the same cited publication).[6] The final dataset which has been used for the RPYS consists of 77,395 papers (from 1989 – 2014) containing 801,571 cited references (from 1900 – 2014). We deliberately decided not to update the dataset of Fiala and Willett (2015) with more recent data: (1) to retain compatibility between theirs and our study, and (2) the year 2014 has a great symbolic value, because it marks the 25th anniversary (a quarter of a century) of the societal changes in Eastern Europe.

In the following, we do not only report the results of the RPYS based on the corresponding dataset. We additionally searched for the times cited information of selected (cited) publications in WoS. We were interested in the general citation impact of citation classics which is not restricted to the investigated RPYS dataset. We also examined the countries and fields of the citing authors which was not possible with the RPYS dataset.

To get a clue how the references cited by Eastern European computer scientists differ from those cited in Western Europe, we inspected the data on computer science papers used in Fiala and Tutoky (2017) and restricted them to the time period 1989-2014 and seven Western European nations: Belgium, Denmark, France, Italy, Netherlands, Sweden, and Switzerland.[7] The final dataset contained 231,775 publication records with 4,676,850 cited references. Although we did not conduct RPYS on the "Western dataset", which was not the goal of this study, we did take a look at the most frequently cited references therein.

[6] Note that the volume and page numbers have been used for the disambiguation process.

[7] We intentionally did not include English-speaking countries like England or Ireland, which could not be fairly compared to non-English-speaking Eastern Europe. Neither did we consider Germany, which spanned over both West Germany and East Germany for some time, nor did we include Spain, which was not a "standard" Western democracy until 1976. The aim was to create a representative group of Western European countries that can be juxtaposed with formerly communist Eastern European nations. It should be considered in the interpretation of the results that only a selected set of Western countries is considered in this study.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

3. Methods

RPYS has been introduced as a method to investigate the historical roots and landmark papers of research fields or topics (Marx et al., 2014). The method is based on investigating cited publications. Most citation analyses – and also the already published scientometric studies focusing on Eastern Europe presented in the following – are based on analyses of the times cited information from Web of Science (by Clarivate Analytics) or Scopus (by Elsevier).

Figure 1 is based on the data downloaded data from WoS. Figure 1 (top) shows the number of cited references (white-dotted line) and the deviation of the number of cited references in one reference publication year from the number of cited references in bordering years (black line). Thus, if the peak is very high for year *t* (and has a positive value for the number of cited references), many cited references (from the computer science field) fall on year *t* – compared to the years [*t*-2, *t*+2]. The RPYS is based on the principle that high peaks (deviations) are hints to important publications in a research field; the peaks in early reference publication years point to the historical roots of the field (citation classics, see bottom chart of Figure 1). It is the premise of the RPYS that important publications for researchers in a field (here: computer science) are often cited in their publications (Bornmann & Marx, 2014). The method is thus based on the wisdom of crowds approach that the judgement of many (here: citation decisions of many authors) contain valid hints about the importance of publications for scientists in a field.

Insert Figure 1 here.

Thor et al. (2018) developed advanced indicators for identifying landmark papers in a field which we used in addition to the basic RPYS functionality. One of the most important indicator is N_TOP10 which can be used to identify highly influential publications. These are publications that belong to the 10% most frequently cited publications over many citing publication years.

4. Results and Discussion

Table 1 shows the publications which are more or less responsible for the highest peaks in Figure 1: 1952, 1965, and 1975. Cited references for the years 2000 and 2002 are not included in the table since there is no publication exceptionally highly cited (the cited references counts are spread relatively equally across many publications). Besides the bibliographic information, the table shows the number of occurrences (in other words, how often the publication has been cited by computer science papers) in absolute and relative terms (relative to the total number of occurrences in one reference publication year).

The selection of the peaks and cited references underlying the peaks for Table 1 are partly arbitrary (by visual inspection of the results in the CRExplorer). Thus, more or less peaks and cited references can be selected for interpreting the historical roots of the field in this study. The final selection of the peaks and cited references should be undertaken by an expert in the field who can interpret the results properly.

Insert Table 1 here.

We will now take a closer look at the publications mentioned in Table 1, which might be interpreted as historical roots for computer science in Eastern Europe. We will describe the topics of the papers, their importance for different scientific fields, their citedness, and (when it markedly differs from the standard pattern of citations from all over the world) also their impact on Eastern European countries. Four of these historical roots stem from 1952. Markowitz's "Portfolio selection" from *The Journal of Finance* received the most citations (n = 54). At first sight it appears unexpected that an influential article was published in a finance journal, but we should be aware of the fact that at the beginning of the 1950s there were hardly any computer science periodicals. The article itself is concerned with theoretical considerations an investor should make to maximize discounted expected returns on investments in marketable securities. It shows that a good investment behavior is to consider

anticipated return as desirable and variance of return as undesirable and argues for a diversified investment portfolio. The underlying techniques used in the article are mathematical probability and statistics as well as geometry.

Although the article itself is not indexed in WoS, there are many references to it therein (7,624 as of 23 January 2019) which can be quickly inspected. Most of them come from the following five subject categories in decreasing order: "Economics", "Business, Finance", "Operations Research & Management Science", "Management", and "Computer Science, Artificial Intelligence". The citation behavior of the publication genuinely corresponds to a typical "sleeping beauty" (a publication with delayed recognition): it lasted 25 years until 1977 to get more than three citations per year. The number of yearly citations then gradually increased but was still less than 100 in 2004 to virtually explode in the years to come and reached hundreds of annual citations from all over the world in the following decade and a half (more than 600 starting with 2016).

The second most-cited article from the same year with 50 citations is Hodgkin's and Huxley's "A quantitative description of membrane current and its application to conduction and excitation in nerve" appearing in *The Journal of Physiology*. It concludes a series of four preceding papers and puts in mathematical form the description of the flow of electric current through the surface membrane of a giant nerve fiber. A look at this article record in WoS reveals that it has been quite steadily cited since 1952 with a continuous growth in citation counts from individual citations immediately after its publication to hundreds of citations annually since 1990 when it was already so well known that it was reprinted in *Bulletin of Mathematical Biology*. The total number of citations achieved so far is 12,748 with about 1% of citations (105) being from publications written in the Russian language, which is second only to English. References to the article are made most frequently from "Neurosciences",

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

"Physiology", "Biophysics", "Mathematical & Computational Biology", and "Multidisciplinary Sciences".

Two further publications from 1952 received 40 citations each (see Table 1): Huffman's article "A method for the construction of minimum-redundancy codes" and Kleene's book "Introduction to Metamathematics", which is a fundamental textbook of mathematical logic. Unlike the former publication, the latter is not indexed in WoS but frequently cited from papers therein. Thus, it requires some effort to merge all the equivalent cited references to get a complete citation picture of this work. The textbook has been consistently moderately cited over the years from fields like "Mathematics", "Computer Science, Theory & Methods", "Logic", "Computer Science, Software Engineering", and "Philosophy" and reaches over 1,500 citations as of January 2019. Among the twenty most frequently citing institutions worldwide there are Moscow State University and Russian Academy of Sciences, which implicates a large impact of this work on the elite institutions of the Russian research system. The article by Huffman is a seminal paper in computer science and information theory introducing the so-called Huffman coding for lossless data compression with many subsequent applications in audio and video file formats, for instance. It has been cited with an increasing trend mostly in "Engineering, Electrical & Electronic", "Computer Science, Information Systems", "Computer Science, Theory & Methods", "Telecommunications", and "Computer Science, Hardware & Architecture" subject categories of WoS only to reach about 2,400 cumulative citations at the beginning of 2019.

The only 1965 article in Table 1 is Zadeh's "Fuzzy sets" published in *Information and Control* with 757 occurrences among the references to the publications from that particular year, which is an almost 20% share. This article laid the foundations of fuzzy sets, which are collections of elements having a varying degree of membership to a set, and fuzzy logic as opposed to Boolean (i.e. yes/no) logic, with applications beyond cybernetics and artificial

intelligence. The paper cites only three references (one of which is Kleene's publication from 1952) but has itself been cited more than 32,000 times in WoS to date. (Since 2007 it has accrued more than 1000 citations every year.) The most frequently citing scientific domains (WoS subject categories) are "Computer Science, Artificial Intelligence", "Engineering, Electrical & Electronic", "Computer Science, Theory & Methods", "Mathematics, Applied", and "Computer Science, Information Systems".

Four further papers in Table 1 were published in 1975. Two of them were authored by Zadeh again (see above): The article "The concept of a linguistic variable and its application to approximate reasoning - I" ranked first with 222 citations and "The concept of a linguistic variable and its application to approximate reasoning - III" ranked fourth with 100 citations, both appearing in *Information Sciences*. These articles are the first and third parts of a series of three papers on the so-called "linguistic variables" whose values are not numbers but words or sentences in a natural or artificial language and which fit well within the framework of fuzzy sets and fuzzy logic. Whereas the first part is mainly an informal introduction into the novel concepts, the third part deals with the calculations of linguistic probabilities. As with Zadeh's 1965 article, there are numerous applications of linguistic variables, primarily in artificial intelligence and cybernetics. The most citations to part I come from "Computer Science, Artificial Intelligence", "Engineering, Electrical & Electronic", "Computer Science, Theory & Methods", "Mathematics, Applied", and "Computer Science, Information Systems". The subject category "Engineering, Electrical & Electronic" ranked fourth instead of second for part III. Both papers have experienced a continuous increase in the number of citations per year achieving hundreds since 2007 (part I) or about a hundred since 2006 (part III). In total, the first part has attracted more than twice as many citations as the second one: around 6,800 compared to 2,700 as of January 2019. Poland, the Czech Republic, and Russia account for more than 5% of the total citations to both papers.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Holland's book "Adaptation in Natural and Artificial Systems" is the second most frequently cited from the year 1975 with 170 citations. It is a fundamental reading on the topic of genetic algorithms (methods inspired by biological systems) with applications in artificial intelligence, game theory, psychology, and economics. As a book it is not present in the WoS core collection, but it is abundantly referred to from the literature indexed therein. To date it has been cited more than 12,500 times, mostly in "Computer Science, Artificial Intelligence", "Engineering, Electrical & Electronic", "Computer Science, Theory & Methods", "Computer Science, Interdisciplinary Applications", and "Operations Research & Management Science" subject categories. Throughout the years it has been only moderately cited until the 1990s, with an explosion after 2000 to reach almost 1000 annual citations from 2012 onwards. About 1% of all citations come from Poland.

The fourth-ranked 1975 publication in Table 1 is Randić's article "On characterization of molecular branching" published in *Journal of the American Chemical Society* with 140 citations. In this paper a well-known topological index was proposed for the description of a chemical structure by a single number reflecting the respective molecular graph. It has found applications beyond mathematical and computer chemistry. It received more than 100 annual citations for the first time in 2003 and has been totally cited over 2,500 times as of January 2019 mainly in "Chemistry, Multidisciplinary", "Computer Science, Interdisciplinary Applications", "Mathematics, Interdisciplinary Applications", "Chemistry, Physical", and "Chemistry, Medicinal" WoS subject categories. Among the top 25 citing countries there is Croatia, Romania, Slovenia, Poland, Yugoslavia, Serbia, Bulgaria, Russia, and Hungary, accounting for more than 20% of all citations. Probably, this large share of "Eastern" citations can surely be partly explained by Milan Randić's affiliation with the University of Zagreb (former Yugoslavia, now Croatia) resulting in many personal relationships to researchers

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

from the region although his 1975 paper was published when the author was a visiting scholar in the USA.

Table 2 lists the publications among the cited references in our dataset (based on the N_TOP10 indicator) which have been cited significantly more frequently than other publications, i.e. in 26 citing years[8] (1989 – 2014). The other publications have been highly cited in less than 26 citing years (or in no citing year at all).

Insert Table 2 here.

Regarding the highly cited publications in Table 2, two of them are Zadeh's papers discussed above. Out of the remaining five publications, only one is a journal article. The other four are books: Harary's "Graph Theory" from 1969, Shafer's "A Mathematical Theory of Evidence" from 1976, and "Computers and Intractability. A Guide to the Theory of NP-Completeness" by Garey and Johnson, and "Introduction to Automata Theory, Languages and Computation" by Hopcroft and Ullman, both from 1979. All four books are fundamental textbooks in their respective subfields of computer science and mathematics[9]. They have been continuously attracting citations in large numbers from the very beginning. Garey's and Johnson's book has been cited the most by far (almost 25,000 citations in WoS), followed by Shafer's publication (about 7,300 citations), Harary's textbook (5,300 citations), and Hopcroft's and Ullman's book (around 4,600 citations).

Hopcroft's and Ullman's book, however, has been the most cited by computer scientists, mainly from the "Theory & Methods", "Software Engineering", "Information Systems", and "Artificial Intelligence" subfields of computer science, with "Mathematics, Applied" as only the fifth most frequently citing research discipline. Similarly with Shafer's book, "Artificial Intelligence", "Theory & Methods", "Information Systems", and

[8] The starting point is the year of the oldest citing paper in the dataset.

[9] The somewhat cryptic title of Shafer's book conceals the introduction of a new theory built upon Bayesian probabilities and statistical inference with applications in artificial intelligence and expert systems.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

"Interdisciplinary Applications" (all in computer science) belong to the top five citing scientific domains, with only "Engineering, Electrical & Electronic" ranked second. In contrast, the other two books receive citations also from other scientific fields to a great extent: "Mathematics", "Mathematics, Applied", "Computer Science, Theory & Methods", "Chemistry, Multidisciplinary", and "Engineering, Electrical & Electronic" (Harary) and "Computer Science, Theory & Methods", "Engineering, Electrical & Electronic", "Operations Research & Management Science", "Mathematics, Applied", and "Computer Science, Information Systems" (Garey and Johnson). Well over 10% of all citations to Harary's work come from Eastern European countries. This share is only about 3% with the other three books.

The remaining publication in Table 2 is the journal article "Optimization by simulated annealing" by Kirkpatrick et al., which appeared in *Science* in 1983. This article introduces a new method of combinatorial optimization inspired by statistical mechanics and presents its application to finding near-optimal solutions (heuristics) for the NP-complete class of problems such as the "traveling salesman", which is generally important to be able to tackle issues that would otherwise be intractable in polynomial time. The impact is far-reaching with effects on the design of algorithms beyond computer science or mathematics, which is documented by the very large number of citations in WoS (about 19,000). More than 1000 citations have been garnered annually since 2015 with most of them coming from "Engineering, Electrical & Electronic", "Computer Science, Artificial Intelligence", "Computer Science, Interdisciplinary Applications", "Computer Science, Theory & Methods", and "Operations Research & Management Science". Citations to this article from Eastern Europe account for less than 1% of the total.

A summary of the top five citing WoS subject categories for all the publications in Tables 1 and 2 is provided in Table 3. Here, the papers are sorted chronologically in

ascending order in 14 table rows and the subject categories are sorted alphabetically in 25 table columns. The publications are denoted by their first author and the year of publication or additionally with the part number as with L. A. Zadeh. Every subject category occurring at least once among the top five citing categories is included in Table 3. We can see that there are many categories with only one occurrence such as "Biophysics" or "Business, Finance". On the other hand, there are a few categories that frequently belong to the main citing sources of the listed papers such as "Computer Science, Theory & Methods" or "Engineering, Electrical & Electronic".

Insert Table 3 here.

In the last step of the statistical analysis, we investigated how the publications cited by Eastern European computer scientists differed from those cited in Western Europe. We did not conduct RPYS on the "Western dataset" but took a look at the most frequently cited references therein. The top 10 cited references are shown in Table 4 and, as we may notice, only some of them are also the most frequently cited by Eastern European researchers. Garey's and Johnson's 1979 book, Zadeh's 1965 article, Holland's 1975 textbook, and the 1983 article by Kirkpatrick et al. belong to these publications. On the other hand, there are a couple of papers which were not among the most frequently cited by Eastern European computer scientists as shown in Table 1 and Table 2. These publications include the 1977 article "Maximum likelihood from incomplete data via the EM algorithm" in *Journal of the Royal Statistical Society* by Dempster et al., Milner's 1989 textbook "Communication and Concurrency", Lowe's 2004 article "Distinctive image features from scale-invariant keypoints" in *International Journal of Computer Vision*, Goldberg's 1989 book "Genetic Algorithms in Search, Optimization and Machine Learning", the 1994 article "A theory of timed automata" by Alur and Dill, and the 1989 article "A tutorial on hidden Markov models

and selected applications in speech recognition" published in *Proceedings of the IEEE* by Rabiner.

Insert Table 4 here.

The paper by Dempster et al. (1977) introduced the two-step expectation-maximization (EM) iterative algorithm for finding maximum likelihood parameter estimates in statistical models dependent on missing data with applications far beyond mathematical statistics and machine learning, computer vision, natural language processing, etc. Milner's (1989) textbook was a continuation of his work on the process calculus (formal language) describing the interaction of two communicating systems, which is useful in software engineering for the design and formal verification of information systems. Lowe's (2004) article is famous for the description of the scale-invariant feature transform (SIFT), a feature selection method used in computer vision for object recognition. In turn, Goldberg's (1989) book is a fundamental tutorial on the theory and application of algorithms based on the mechanics of natural selection and genetics. The purely theoretical journal article by Alur and Dill (1994) introduced the invention of timed finite state machines, which has been used in the design and study of real-time systems since then. The well-known tutorial by Rabiner (1989) is actually an overview of the theoretical aspects of Markovian statistical modeling methods including a few examples of applications in speech recognition when language is modeled as a signal.

At first sight one can immediately notice that the cited references in Table 4 are more recent than those in Tables 1 and 2. A half of them are publications from the late 1980s, 1990s, and even as recent as 2004, which is something unseen with the top references cited by Eastern European researchers. Therefore, it would appear that computer scientists from Eastern Europe are more conservative and tend to cite older well-known research compared to

their Western European counterparts who do not shy away from building upon the last breakthrough results in science.

5. Conclusions

RPYS is a method of inspecting cited references in a collection of bibliographic records. The goal is to identify those references that are outstanding in terms of the total number of citations. Very influential publications are identified by the difference between its number of citations and the number of citations of other publications from the same time period. The RPYS method has been developed to unveil the historical roots of a scientific discipline, research topic, scientific journal, or a single researcher or group of researchers.

In this study, we used the CRExplorer software to examine around 800,000 references cited in approximately 80,000 computer science papers, which were (co-)authored by Eastern European researchers, published between 1989 and 2014 and indexed in the WoS database. This dataset is basically the same that was utilized in Fiala and Willett (2015). The authors explored the evolution of Eastern European computer science after the fall of the Iron Curtain; however, no attempt was made to analyze the cited references of the literature. Thus, we have analyzed in this follow-up study what the landmark papers of Eastern European researchers are. Is there a specific Eastern European pattern in the use of the literature?

In the first step of the analysis, we determined three peak years, in which publications were cited more frequently than publications in surrounding years: 1952, 1965, and 1975. We identified the highly cited works published in these years (responsible for the peaks) and discussed their content and impact on various scientific fields (using further WoS data). We also inspected the publications which belonged to the most frequently cited publications over the 26 citing years period under study. Additionally, for a first comparison of the differences in the references cited in Eastern Europe and in the West, we investigated almost five million cited references from seven Western European countries.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Our results show that the works that seem to have had the most impact on computer science research in Eastern European countries after the break-up of the former communist bloc are Markowitz's "Portfolio selection" from 1952, Zadeh's "Fuzzy sets" from 1965 and "The concept of a linguistic variable and its application to approximate reasoning – I" from 1975. The analyses also pointed to a few classic textbooks like Shafer's "A Mathematical Theory of Evidence" from 1976, Garey's and Johnson's 1979 book "Computers and Intractability. A Guide to the Theory of NP-Completeness", Harary's "Graph Theory" from 1969, and Hopcroft's and Ullman's 1979 textbook "Introduction to Automata Theory, Languages and Computation". Furthermore, our results reveal that computer scientists from Eastern Europe are more conservative in their citation behavior and tend to refer to older and more established research than their counterparts from the West.

In our future work, we would like to further verify our preliminary finding that the computer science classics in Western Europe are somewhat less conservative than those in Eastern Europe.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

References

Allik, J. (2013), "Factors affecting bibliometric indicators of scientific quality", *Trames*, Vol. 17 No. 3, pp. 199-214.

Bornmann, L. and Marx, W. (2014), "The wisdom of citing scientists", *Journal of the American Society of Information Science and Technology*, Vol. 65 No. 6, pp. 1288-1292.

Fiala, D. and Tutoky, G. (2017), "Computer science papers in Web of Science: A bibliometric analysis", *Publications*, Vol. 5 No. 4, art. no. 23.

Fiala, D. and Willett, P. (2015), "Computer science in Eastern Europe 1989-2014: a bibliometric study", *ASLIB Journal of Information Management*, Vol. 67 No. 5, pp. 526-541.

Grančay, M., Vveinhardt, J. and Šumilo, Ē. (2017), "Publish or perish: How central and eastern European economists have dealt with the ever-increasing academic publishing requirements 2000–2015", *Scientometrics*, Vol. 111 No. 3, pp. 1813-1837.

Ivanovic, D. and Ho, Y.-S. (2014), "Independent publications from Serbia in the science citation index expanded: a bibliometric analysis", *Scientometrics*, Vol. 101 No. 1, pp. 603-622.

Jurajda, Š., Kozubek, S., Münich, D. and Škoda, S. (2017), "Scientific publication performance in post-communist countries: Still lagging far behind", *Scientometrics*, Vol. 112 No. 1, pp. 315-328.

Karamourzov, R. (2012), "The development trends of science in the CIS countries on the basis of some scientometric indicators", *Scientometrics*, Vol. 91 No. 1, pp. 1-14.

Kozak, M., Bornmann, L. and Leydesdorff, L. (2015), "How have the Eastern European countries of the former Warsaw pact developed since 1990? A bibliometric study", *Scientometrics*, Vol. 102 No. 2, pp. 1101-1117.

Lovakov, A. and Agadullina, E. (2017), "Bibliometric analysis of publications from post-Soviet countries in psychological journals in 1992-2016", in *Proceedings of the 16th International Conference on Scientometrics and Informetrics (ISSI 2017)*, Wuhan, China, pp. 125-135.

Marx, W., Bornmann, L., Barth, A., & Leydesdorff, L. (2014). Detecting the historical roots of research fields by reference publication year spectroscopy (RPYS). *Journal of the Association for Information Science and Technology*, 65(4), 751-764.

Popovic, A., Antonic, S. and Matutinovic, S. F. (2012), "Rapid changes of Serbian scientific journals: their quality, visibility and role in science locally and globally", *Communications in Computer and Information Science*, Vol. 317, pp. 61-70.

Prathap, G. (2018), "Performance of research universities in post-communist countries", *Scientometrics*, Vol. 117 No. 3, pp. 2037-2039.

Radosevic, S. and Yoruk, E. (2014), "Are there global shifts in the world science base? Analysing the catching up and falling behind of world regions", *Scientometrics*, Vol. 101 No. 3, pp. 1897-1924.

Teodorescu, D. and Andrei, T. (2011), "The growth of international collaboration in East European scholarly communities: a bibliometric analysis of journal articles published between 1989 and 2009", *Scientometrics*, Vol. 89 No. 2, pp. 711-722.

Thor, A., Bornmann, L., Marx, W. and Mutz, R. (2018), "Identifying single influential publications in a research field: New analysis opportunities of the CRExplorer", *Scientometrics*, Vol. 116 No. 1, pp. 591–608.

Thor, A., Marx, W., Leydesdorff, L. and Bornmann, L. (2016a), "Introducing CitedReferencesExplorer (CRExplorer): A program for Reference Publication Year

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Spectroscopy with Cited References Standardization", *Journal of Informetrics*, Vol. 10 No. 2, pp. 503-515.

Thor, A., Marx, W., Leydesdorff, L. and Bornmann, L. (2016b), "New features of CitedReferencesExplorer (CRExplorer)", *Scientometrics*, Vol. 109 No. 3, pp. 2049-2051.

Vanecek, J. (2008), "Bibliometric analysis of the Czech research publications from 1994 to 2005", *Scientometrics*, Vol. 77 No. 2, pp. 345-360.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Table 1. Cited references with the largest number of occurrences in reference publication years with the highest peaks in Figure 1.

Cited reference	Number of occurrences (absolute)	Number of occurrences (in percent)
1952		
Markowitz, H. (1952). Portfolio selection. *The Journal of Finance*, 7(1), 77-91.	54	5.7
Hodgkin, A. L., & Huxley, A. F. (1952). A quantitative description of membrane current and its application to conduction and excitation in nerve. *The Journal of Physiology*, 117(4), 500-544.	50	5.3
Huffman, D. A. (1952). A method for the construction of minimum-redundancy codes. *Proceedings of the IRE*, 40(9), 1098-1101.	40	4.2
Kleene, S. C. (1952). *Introduction to Metamathematics*. North Holland/Van Nostrand, Amsterdam/New York.	40	4.2
1965		
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.	757	19.2
1975		
Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning - I. *Information Sciences*, 8(3), 199-249.	222	3
Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.	170	2.3
Randić, M. (1975). On characterization of molecular branching. *Journal of the American Chemical Society*, 97(23), 6609-6615.	140	1.9
Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning - III. *Information Sciences*, 9(1), 43-80.	100	1.4

Note. The percentages are calculated on the basis of the total number of cited references in one year

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Table 2. Cited references which belong to the 10% most frequently cited in more citing years than other cited references. The table shows the cited publications which are highly cited over 26 citing years.

Cited reference	Number of citing years	Number of citations
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.	26	757
Harary, F. (1969). *Graph Theory*. Addison-Wesley, Boston.	26	105
Zadeh, L. A. (1975). The concept of a linguistic variable and its application to approximate reasoning-I. *Information Sciences*, 8(3), 199-249.	26	222
Shafer, G. (1976). *A Mathematical Theory of Evidence*. Princeton University Press, Princeton.	26	189
Garey, M. R., & Johnson, D. S. (1979). *Computers and Intractability. A Guide to the Theory of NP-Completeness*. W.H., Freeman and Company, New York.	26	514
Hopcroft, J. E., & Ullman, J. D. (1979). *Introduction to Automata Theory, Languages and Computation*. Addison-Wesley, Boston.	26	181
Kirkpatrick, S., Gelatt Jr., C.D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680.	26	199

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".

Table 3. Highly cited papers and the ranks of their top five citing Web of Science categories (1 is best, 5 is worst).

	Biophysics	Business, Finance	Computer Science, Artificial Intelligence	Computer Science, Hardware & Architecture	Computer Science, Information Systems	Computer Science, Interdisciplinary Applications	Computer Science, Software Engineering	Computer Science, Theory & Methods	Economics	Engineering, Electrical & Electronic	Chemistry, Medicinal	Chemistry, Multidisciplinary	Chemistry, Physical	Logic	Management	Mathematical & Computational Biology	Mathematics	Mathematics, Applied	Mathematics, Interdisciplinary Applications	Multidisciplinary Sciences	Neurosciences	Operations Research & Management Science	Philosophy	Physiology	Telecommunications
Hodgkin (1952)	3															4				5	1			2	
Huffman (1952)				5	2			3		1															4
Kleene (1952)							4	2						3			1						5		
Markowitz(1952)		2	5						1						4							3			
Zadeh (1965)			1		5			3		2								4							
Harary (1969)								3		5		4					1	2							
Holland (1975)			1			4		3		2												5			
Randić (1975)						2					5	1	4						3						
Zadeh (1975 - I)			1		5			3		2								4							
Zadeh (1975 - III)			1		5			2		4								3							
Shafer (1976)			1		4	5		3		2															
Garey (1979)					5			1		2								4				3			
Hopcroft (1979)			4		3		2	1										5							
Kirkpatrick (1983)			2			3		4		1												5			

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. *Aslib Journal of Information Management* (in press). DOI: 10.1108/AJIM-06-2019-0142".

Table 4. The ten most frequently cited references in computer science papers by Western European researchers.

Cited reference	Number of citations
Garey, M. R., & Johnson, D. S. (1979) *Computers and Intractability. A Guide to the Theory of NP-Completeness*. W.H., Freeman and Company, New York.	1,663
Zadeh, L. A. (1965). Fuzzy sets. *Information and Control*, 8(3), 338-353.	1,011
Dempster, A. P., Laird, N. M., & Rubin, D. B. (1977). Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society. Series B (Methodological).* 39(1), 1-38.	954
Milner, R. (1989). *Communication and Concurrency*. Prentice Hall, Upper Saddle River, NJ, USA.	942
Lowe, D. G. (2004). Distinctive image features from scale-invariant keypoints. *International Journal of Computer Vision*, 60(2), 91-110.	935
Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. Addison-Wesley, Boston.	846
Alur, R., & Dill, D. L. (1994). A theory of timed automata. *Theoretical Computer Science*, 126(2), 183-235.	636
Holland, J. H. (1975). *Adaptation in Natural and Artificial Systems*. University of Michigan Press, Ann Arbor.	573
Rabiner, L. R. (1989). A tutorial on hidden Markov models and selected applications in speech recognition. *Proceedings of the IEEE*, 77(2), 257-286.	534
Kirkpatrick, S., Gelatt Jr., C.D., & Vecchi, M. P. (1983). Optimization by simulated annealing. *Science*, 220(4598), 671-680.	532

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. *Aslib Journal of Information Management* (in press). DOI: 10.1108/AJIM-06-2019-0142".

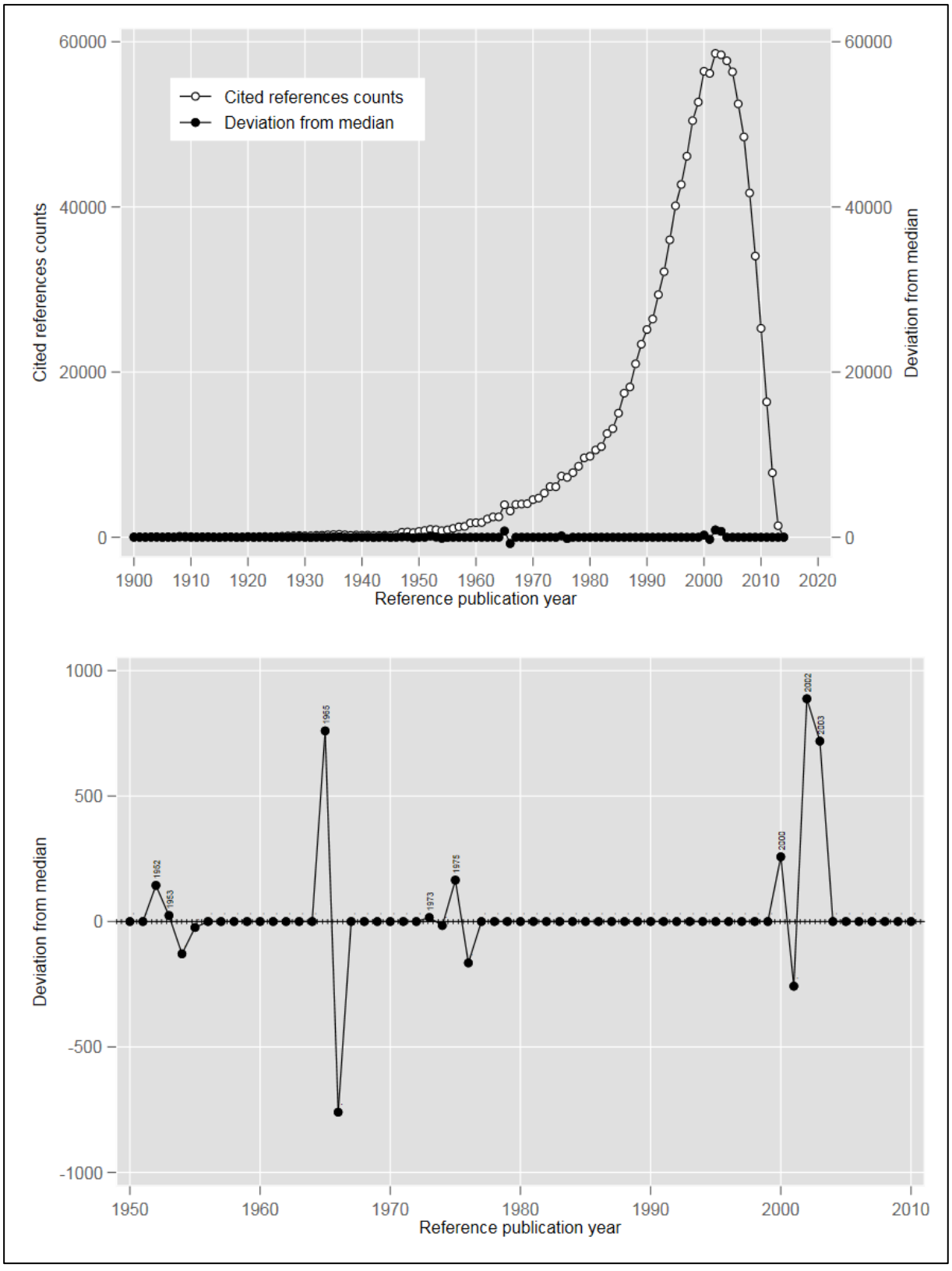

Figure 1. The top chart shows the number of cited references (white-dotted line) and median deviations of cited references (black line). The peaks (with positive values) in the black line show reference publication years with a significantly greater number of cited references than bordering years. The bottom chart focuses on the peak years.

Preprint of "Fiala, D., & Bornmann, L. (2020). Reference publication year spectroscopy (RPYS) of computer science papers from Eastern Europe. ***Aslib Journal of Information Management*** (in press). DOI: 10.1108/AJIM-06-2019-0142".