

Authorship and Content Analysis of Engineering Education Research: A Case Study

<http://dx.doi.org/10.3991/ijep.v6i2.5577>

Sadia Nawaz^{1,2}, Johannes Strobel³

¹ Purdue University, West Lafayette, IN, USA

² University of Melbourne, Melbourne, Australia

³ University of Missouri, Columbia, MO, USA

Abstract—In this paper, *Engineering Education* as a discipline has been analyzed by taking *IEEE Transactions on Education (IEEE T Educ)* as a case; for examining the various trends that have been emerging over time. Based on various criteria of authorship and citation, an effort is made to highlight the main contributors or top authors of this engineering education community. It was found that authorship trends have been shifting more towards collaboration. It was also found that the authorship community is growing, both in terms of publications and publishing authors. Study of citation patterns during the last decade, reveals a high citation count per article, which indicates a high readership of this journal. Later, the study of authorship and citation patterns shed light on the trend that multi-author articles are cited more often than single-author articles. This study was compared with earlier studies in the field of Engineering Education Research (EER) using *keyword analysis and temporal evolution and distribution of keywords*. Additionally key-phrase and topic modeling was performed to identify leading and evolving research areas within the EER., Analysis of word co-occurrence was performed to discover the main context in which the keywords have been used. Lastly, topic modeling techniques were applied for probabilistic distributions of *IEEE* topics and the results were in line with earlier studies.

Index Terms—bibliometric analysis, citation analysis, collaboration, engineering education research, learning analytics, social network analysis

I. INTRODUCTION AND MOTIVATION

RESEARCH ON EDUCATIONAL ISSUES in engineering, computer science, and other technical areas is growing fast [1] due to the increased interest of funding agencies [2], the fact that facilitation of cross disciplinary research has attracted substantial attention in recent years [3], also, a growing number of scholars dedicating their career to the intersection of technical fields and education[4], and lastly, higher demands to show the effectiveness of programs for student achievement and retention. At the same time, fields such as engineering education research and computer science education research are solidifying and developing their own identity [5-6], which is strongly indicated by the maturity and growth of field-specific conferences such as FIE (Frontiers in Education) and EDUCON, as well as the new establishment of a special interest group (at the American Educational Research Association), a research interest group (within the National Association for Research in Science Teaching), and research focused on engineering/computer science education within technical and general education communities.

The body of literature in engineering education (EE) is growing substantially and reaching a point where a detailed historical analysis could provide insights and moments for reflection. While such analysis is useful and already conducted for the *Journal of Engineering Education (JEE)* [7-8], for the *International Journal of Electrical Engineering Education (IJEEE)* [9] and on a larger scale for the entire field of EE [10]; meaningful analysis of a journal such as *IEEE T Educ* carries numerous additional benefits: (1) Results might be used to recognize key topics and actors in the community. (2) Recognition of key authors within *IEEE* can provide a link to those who have expertise in ‘advanced pedagogical methods’ [11] (3) A historical analysis of how certain topics developed and possibly faded could provide insights on how the journal influenced and shaped the field, and could support the development of a shared narrative and memory of the field. (4) Such analysis can further aid in providing a ‘map of the terrain’ for novice researchers in the field [12-13]. (5) An analysis of the content, authorship patterns, and impact of articles could lead to the development of benchmarks and strategic planning. As P. Wankat [11], suggests that the EE journals should devise plans for further enhancement of this field. (6) The journal *IEEE T Educ* started in its current form in 1963 with aims “both scientific and educational, grounded in the theory and practice of electrical and computer engineering” [14], making it one of the oldest scholarly publications in the intersection of technical fields and education. Thus, analyzing *IEEE T Educ* provides not only insights into the journal itself, but also a historical window into the history of the emerging discipline of EE and EER.

This paper aims to address the following questions: Who are the main contributors to *IEEE*? How are the authorship trends changing over time? Can citation be associated with collaboration? What themes have been explored in the past, and what are the emerging themes in this journal? What have been the topics of highly cited publications?

II. DATA COLLECTION AND PROCESSING

In recent years bibliometric analysis of publications has been gaining boom in various disciplines including EER [7, 9, 15-16]. Currently, citation data is available from various sources such as: Institute for Scientific Information’s (ISI) citation databases, Scopus and Google Scholar. Some studies suggest the usage of multiple databases for research assessment [17-18]. However, others suggest that the accuracy of Google Scholar still needs to be tested [19-20] and that the results of ISI and Scopus are comparable [21-22]. Some studies further highlight the

advantages associated with ISI's Web of Science® (WoS) in terms of processing times[17] and in terms of the available information [23-25]. Also, WoS provides access to the world's leading citation databases, over 12,000 of the highest impact journals and over 150,000 conference proceedings; it is widely used by researchers, scholars, administrators, faculty and students. The coverage in WoS dates back to 1900 and the information spans over 250 disciplines. Therefore, the data for this analysis has been taken from the ISI's WoS. WoS as selected database for the current study makes it possible to explore scientific collaboration of interdisciplinary research as well as to define a field of inquiry by identifying citation patterns and core publications [26]. The data for current study¹ includes all the *IEEE T Educ* articles and proceedings from 1963 to 2011. The authors manually verified the data to incorporate name variants and missing keyword fields (the bibliographic entry for some of the articles in WoS showed empty keyword field, where as it was present in the article itself). Descriptive statistics for this data are provided next: a) in total there are 2250 articles out of which only twelve (12) are proceedings papers. b) There have been contributions to this journal from 69 different countries. The countries with highest publication counts in *IEEE T Educ* include USA (54.58%), Spain (5.87%), UK (4.44%), Canada (4.27%) and Australia (2.18%). c) The institutes with most number of articles in this journal are: Georgia Institute of Technology (49), US Department of Defense (49), California State University System (45), University of California System (45) and Florida State University System (37). Since the *IEEE T Educ* dataset spans 49 years so it has been divided into decades; except for the last time span which is only 9 years, i.e. 2003–2011 (for the sake of convenience, this time frame will also be called a decade). Before the *IEEE* was formed from the merger of two professional societies, the predecessor to the *IEEE T Educ* was the *Institute of Radio Engineers (IRE) Transactions on Education* [27]. However, it was decided to study the journal in its current format and therefore only the issues from 1963 onwards were considered. Due to space limitations, the analysis of the journal will be done comprehensively for the last time frame (2003–2011) along with a thorough explanation of the adopted methodology. However, for earlier decades, a combined analysis will be presented for comparison and investigation.

III. AUTHORSHIP ANALYSIS

In this section, the main contributors during the last decade of this journal are analyzed. Contributors within this community are regarded as leading authors. Studies on leading authors and their collaboration date back to Derek Price and other early bibliometricians [28]. In this paper, contributors are defined by many standards: (1) First classification is based on whether the article is single-authored or multi-authored. (2a) Further classification is then based on each type e.g., in the case of multi-authored articles, an author pair is brought out as a 'main contributor' based on the number of joint articles written by the two authors. Similarly, for a given author of a multi-authored article, 'diversity' is defined as the number of other authors he or she might have worked with. (2b) For single authors, this level of classification brings out only those authors who have the highest publication counts. (3) Later, the authors

of single- and multi-authored articles are classified together, based on the 'citation counts' of the articles. Regarding such classification, Klavans and Boyack [29] have suggested that individual publication and citation counts can be taken as an indicator for identifying the leading researchers within a field. Similarly, other researchers suggest that citation and publication counts are obvious indicators of productive and creative scientists within a field [30-32]. These prolific researchers have more impact on the field, and the likelihood that these papers are selected as "creative work" is higher [33]. Another method for identifying the main researchers within a field is through peer evaluation [34]. However, bibliometrics-based methods are easier and more efficient, especially due to the availability of research databases such as Web of Science and Scopus. [33].

Since *IEEE T Educ* is an interdisciplinary journal and its scope covers "education research, methods, materials, programs, and technology in electrical engineering, computer engineering, and fields within the scope of interest of *IEEE*" [14], it can thus be suggested that the main contributors are authors "who exert influence in both their own discipline and other disciplines". These researchers can quickly adapt to new knowledge and ideas from related disciplines and for these authors such inter-disciplinary results in creative problem solving [34].

A. Who are the main contributors to this journal?

Before answering this question, it is necessary to define the following terms as used in this paper: *Node* or *vertex* in this study represents an author; *edge* exists between two authors if they are co-authors of an article; *unique edge count* will represent the overall count of co-authors (irrespective of their mutual article or publication count). Thus, the co-authorship is represented by a line between two different authors or by a loop on a node (as shown in figure 4). Detailed definition of these and various other graph theory related terms, is available at [35-38] The following analysis has been performed with NodeXL tool – an open source template for Microsoft® Excel® [39] which allows users to work on different worksheets such as: 'Edges' worksheet can be used to compute the inter/intra collaboration; 'Vertices' worksheet allows the display and computation of individual node properties such as degree, betweenness and centrality etc.. Table I provides a summary of various statistics, for articles published in the last decade. This table shows that out of the 609 published articles: 129 are single- and the remaining 480 are multi-authored articles. In total, there have been contributions from 1636 different researchers. The table entry for "single-vertex connected component"² tells that out of the 129 single-authored publications, 109 are by those authors who do not have any collaborative work within this journal (during the mentioned time-frame).

1) Multi-author trends:

The authorship trends during the last decade have been presented in Figure 1. Here the authors have been grouped and colored (group G1 – G13) based on the count of distinct researchers they have worked with³. The dots repre-

²,* A graph is connected if a path exists between all pairs of vertices. If the graph is not connected then it can be divided into connected components where each component is mutually exclusive with every other component.

³ The complete description about the chosen layout and graph metrics is provided in the supplementary material.

¹ Most data was downloaded in early 2012, unless specified otherwise.

sent the authors and the collaboration between authors is represented by lines or edges. The darker this line gets, the greater the number of articles between those two authors. From this figure it can be seen that the degree of authors ranges from 1 – 23; showing that some authors have worked with just one other author while some have worked with 23 other authors.

From Figure 1 some of the outliers are apparent such as in G10 there are 24 authors each with degree 23; it was found that they all co-authored in one article “The Olin curriculum: Thinking toward the future”. Similarly, it was found that there are 35 authors each with degree 10. Since degree in this context is a measure of collaboration with distinct other researchers; so based on this information it was decided to bring out the most prominent authors of multi-authored articles. However, before processing any further, it was necessary to eliminate the outliers (such as, the one with a very high count of co-authors - 24); since they could skew the results of the analysis and may bring the authors of such articles on top (where as the aim is to study the general authorship trends and not just one article where a larger number of co-authors appeared). For this purpose the article–author distribution was extracted as shown in Figure 2. It was found that 21% of articles are single-authored and that the remaining 79% -the co-authored articles, are distributed as follows: 22.7% are a result of collaboration between two authors, 25.3% are a result of collaboration between three authors, and the remaining 31% are the collaborative work of four or more co-authors. It was found that only 18 articles (nearly 3% of all articles) were a result of collaboration between seven (7) or more authors. Therefore, these articles were considered outliers and were excluded from the analysis.

Thus, the main contributors based on degrees of diversity are presented in Table II, which lists the top ten authors with highest degree counts and their associated publication and citation counts⁴.

The above data was retrieved manually from WoS to include author name variants etc. These top authors were further investigated to study their authorship patterns in terms of collaboration. Collaboration as defined in Oxford dictionary [40] is the ‘action of working with someone to produce something’ and in current context it represents co-authorship of an article by two or more researchers. (This term can be extended to institutes and even to countries and hence extended collaboration patterns can be extracted - but due to space limitation, this study cannot be presented here). Smith [41] was one of the early researchers who suggested that multi-author papers can be used ‘as a proxy measure for collaboration among groups of researchers’. However, he reported that such studies only provide an ‘approximate measure of group efforts’ [42-43]. Despite this limitation, many studies have utilized co-authorship as a measure of collaboration [44-47] and the general consensus is that the growth in multiple-authorship is a proof of increase in collaboration [45-49]. Therefore, the collaboration patterns of top authors of Table II were further examined and it was found that the total number of authors brought out by these top ten researchers is 54 and none of these top authors had any single-author publication (during 2003-2011 in this journal). Moreover, all of these authors are from Spain except Ramachandran, RP and Garcia, A and they both are affiliated with the ECE departments in

TABLE I.
AUTHOR - PUBLICATION SUMMARY FOR *IEEE T EDUC* (2003–2011)

Total articles	609
Total unique authors	1636
Unique links between authors (This link is represented by a loop in case of single-author articles and straight line otherwise)	110
Multi-author articles	480
Single-author article count (self-loops)	129
Authors forming co-authorship cluster (Connected components*)	507
Number of authors of single-author articles who have never collaborated (single-vertex connected components)	109
Maximum authors in a cluster (Connected component)	24
Maximum links in a cluster (Connected Component)	552

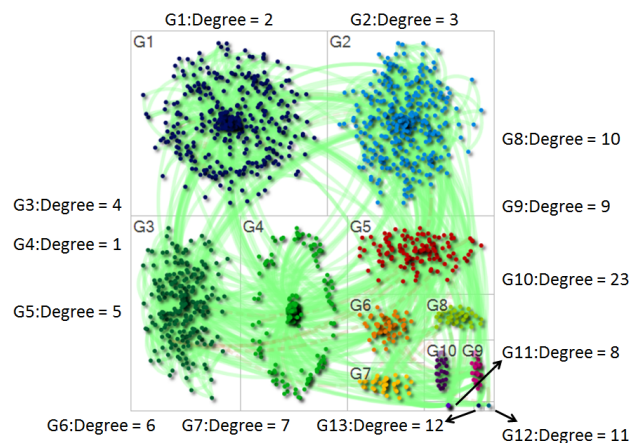


Figure 1. Overall authorship patterns during 2003-2011.

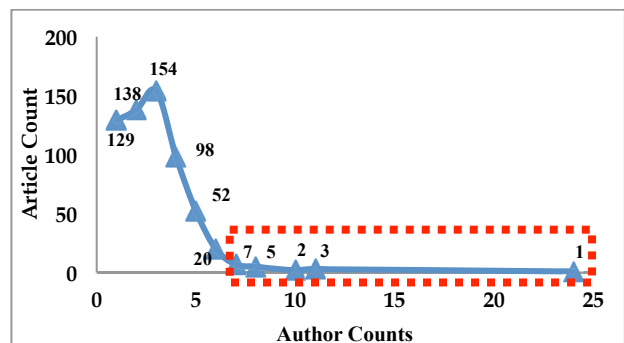


Figure 2. Article-author distribution for the of study collaboration trends

TABLE II.
TOP TEN CONTRIBUTORS OF MULTI-AUTHOR ARTICLES BASED ON DEGREE/DIVERSITY, 2003-2011

Authors	Degree	Publications	Citations
Martinez, M	9	2	15
Ramachandran, RP	9	3	15
Rodriguez, A	8	3	5
Martinez-Torres, MR	8	3	28
Soria, E	8	2	7
Magdalena, R	7	2	20
Rodriguez, S	7	3	19
Garcia, A	7	3	19
Kloos, CD	7	4	10
Sanchez, FM	7	3	28

⁴ Citation data was downloaded from WoS in October, 2013

USA. Sanchez, FM and Rodriguez, S. belong to the faculty of CS; Kloos, CD is affiliated with Telematic Engineering and Martinez-Torres, MR is affiliated with Business and Management School and all the others belong to EE departments with research focus in communication systems and digital signal processing etc.

As mentioned earlier, the main contributors of multi-authored articles can also be defined by joint publications between two authors. Therefore, the analysis of such articles is presented in Table III. Here, the presence of same co-author(s) in more (than one) papers indicates (strong) collaboration and hence suggests that these authors share the same interests. Interestingly, all of the top author-pairs for this journal during 2003-2011 have jointly worked in three (3) articles. Moreover, Chang, GW; Chang, HM and Yeh, ZM worked together in three articles. These main contributors belong to USA, Taiwan and Spain and disciplinary affiliations include not only EE, CS, Telematic Engineering but also Mechatronics.

Lastly, the multi-author article trend was studied based on the counts of unique authorship pairs. Therefore, during the last decade only six author pairs have co-authored three (3) articles, eighty five (85) author pairs co-authored two (2) articles and 2,638 author pairs have worked on just one article together⁵. Thus it can be concluded that co-authorship between same author pairs is not very common in this community. Therefore, the total count of unique author-pairs is high:

$$3(6) + 2(85) + 1(2638) = 2826.$$

2) Single-author trends:

Next, the single-authorship patterns were studied using NodeXL in which 'self-loop count' with value zero (0) shows multi-authored patterns and positive value (greater than or equal to one) indicates single-authored articles. It was found that out of 119 single authored articles, 109 are by the authors with no collaboration within *IEEE T Educ*. And only 10 of the single authored articles are by the authors who have some collaborative work in *IEEE T Educ* during the last decade⁶. The summary is: Hwang, GJ; Chen, WF; Shirsavar, SA and Aleman, JLF have worked with 2 other persons (in addition to having single authored articles) and Ruiz-del-Solar, J; Frolik, J; Parent, DW; Sarkar, NI; Hamblen, JO and Mazhari, B have worked with one other person in addition to their single authored work. Lastly, none of the remaining 109 authors have worked with any other researcher in *IEEE T Educ* from 2003-2011. Thus, the patterns of single-authored articles can be summed up as: out of the 129 single-authored articles, 85.3% were written by authors who have only one single-authored publication; the remaining 14.73% were written by authors who have two or three single-authored publications. Based on these publication counts, the top authors of single authored articles were extracted as listed in Table IV.

This table presents relatively diverse authorship community in that the authors belong to countries such as New Zealand, India, Korea, Romania, Israel, USA and Spain. Interestingly Buiu, C belongs to Department of Automatic

⁵ This information is based on the data preparation for input into NodeXL, and its complete explanation is provided in Supplementary Material.

⁶ NodeXL counts the single authored publications as a work with distinct researcher. For further details please see "NodeXL degree calculation for self-loops" as provided in the Supplementary Material.

TABLE III.
MAIN CONTRIBUTORS OF MULTI-AUTHOR ARTICLES DURING 2003–2011 BASED ON COLLABORATION

Author Names	Author Names	Article Count
Chang, GW	Yeh, ZM; Chang, HM	3
Yeh, ZM	Chang, HM	3
Garcia, A	Rodriguez, S	3
Munoz-Organero, M	Kloos, CD	3
Schubert, TF	Kim, EM	3

TABLE IV.
MAIN CONTRIBUTORS OF SINGLE-AUTHOR ARTICLES BASED ON PUBLICATION COUNT, 2003-2011

Authors	Article Counts	Authors	Article Counts
Abramovitz, A	3	Smaill, CR	2
Mazhari, B	2	Roy, SCD	2
Cavicchi, TJ	2	Lim, DJ	2
Lopez-Martin, AJ	2	Buiu, C	2
Rothwell, EJ	2	-	-

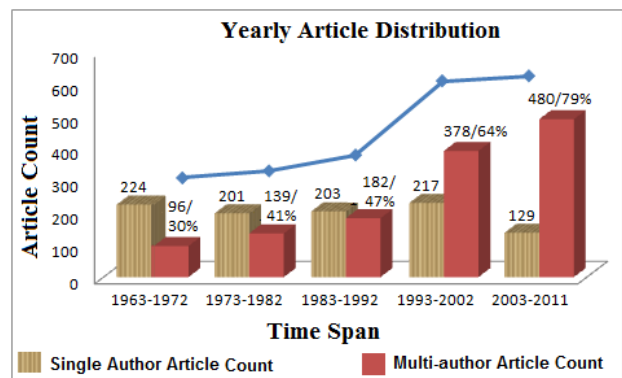


Figure 3. Article distribution per decade for *IEEE T Educ*, 1963–2011

Control and Systems Engineering and all the other authors belong to Electrical and Computer Engineering Departments. This completes the authorship analysis for the last decade. Next, similar analysis will be done for earlier decades, and their trends will be compared and contrasted to see the emergence of the journal throughout its life span.

B. How the authorship trends are changing over time?

Figure 3 presents various authorship trends. Here the upper curve shows the overall article count per decade for this journal; it clearly shows that the journal itself is growing. The lower bar graphs show the distribution of articles into single- and multi-authorship. It is evident that the general trend for authorship is shifting towards collaboration and it is a general trend followed these days [50-52]. However, in earlier decades, inclination was more towards single authorship; 70% of articles in the very first decade were single-authored, which dropped gradually to 21% in the later years (2003–2011).

The author-publication summary (similar to Table I) is presented next in Table V; it spans from 1963 to 2011. Growth of the journal is again evident from: total article count and also from the community of authors within this journal. In the early years (1963–1972) there were contributions from only 397 different authors, but in the last time frame (2003-2011) this number reached 1,636 dis-

PAPER
AUTHORSHIP AND CONTENT ANALYSIS OF ENGINEERING EDUCATION RESEARCH: A CASE STUDY

TABLE V.
OVERALL AUTHOR - PUBLICATION SUMMARY FOR *IEEE T EDUC* (1963–2011)

Period	1963 –1972	1973–1982	1983–1992	1993–2002	2003–2011
Total articles	320	340	385	595	609
Total unique authors (vertices)	397	530	615	1258	1636
Unique links (edges)	170	178	171	188	110
Multi-author articles	96	139	182	378	480
Single-author articles (self-loops)	224	201	203	217	129
Author clusters based on co-authorship (connected compnt)	265	304	328	536	507
Authors of single-authored articles who never collaborated	181	175	171	189	109
Maximum authors in a cluster (nodes in a connected compnt)	8	14	11	28	24
Maximum links in a cluster (edges in a connected compnt)	45	85	110	756	552

TABLE VI.
MAIN CONTRIBUTORS BASED ON CITATION COUNTS, 2003-2011

No. of Cites	Authors	Year	Cites year	Title
51	Sanchez, J; Dormido, S; Pastor, R; Morilla, F	2004	6.38	A Java/Matlab-based environment for remote control system laboratories: Illustrated with an inverted pendulum
50	Graesser, AC; Chipman, P; Haynes, BC; Olney, A	2005	7.14	AutoTutor: An intelligent tutoring system with mixed-initiative dialogue
48	Casini, M; Prattichizzo, D; Vicino, A	2003	5.33	The automatic control telelab: A user-friendly interface for distance learning
48	Gillet, D; Ngoc, AVN; Rekik, Y	2005	6.86	Collaborative web-based experimentation in flexible engineering education
41	Tzafestas, CS; Palaiologou, N; Alifragis, M	2006	6.83	Virtual and remote robotic laboratory: Comparative experimental evaluation
37	Toh, BY; Cahill, R; Fusco, VF	2003	4.11	Understanding and measuring circular polarization
27	Guzman, E; Conejo, R	2005	3.86	Self-assessment in a feasible, adaptive web-based testing system
27	Leva, A	2003	3.0	A hands-on experimental laboratory for undergraduate courses in automatic control
25	Hurley, WG; Lee, CK	2005	3.57	Development, implementation, and assessment of a web-based power electronics laboratory
24	Lindsay, ED; Good, MC	2005	3.43	Effects of laboratory access modes upon learning outcomes

tinct authors. Thus, author's community is also growing. In this table, the authors of single-authored articles with no collaborative work within the journal are distributed as: during the first decade, 80.8% of authors of single-authored articles never collaborated (within this journal). During the second decade, this percentage increased to 87.1%, and in the third decade, it decreased to 84.24%. During the fourth and fifth decades, this percentage was 87% and 84.5%, respectively. So, on average, only 15% of the authors of single-authored articles have also been involved in collaboration (within the community of this journal, since 1963). This percentage is computed by taking the ratio of the single-vertex connected component counts and the self-loop counts. In Figure 3, it was suggested that the authorship trends in this journal are gradually shifting towards multi-authorship. Another evidence of this trend is seen from Table V – the 'count of connected components,' which was 265 in the first decade and 507 in the last decade. So the authors of this journal are forming a close-knit community as more and more get connected in clusters of co- authorship. It is also interesting to observe that not only the count of connected components is increasing, but also the count of co-authors (maximum nodes in a connected component) is increasing. In the first decade, this count was eight (8) and it rose to twenty four (24) during the last decade. The graphical display of overall authorship patterns from 1963 to 2011 (per decade) are shown in Figure 4.

The authorship graphs have been split into single- and multi-authorship (separated by arrows). Here authors have been placed in a circular layout (rather than the Harel-Koren Fast Multiscale layout which was used in Figure 1) to better display the changes in authorship patterns within the community of *IEEE*. Left side of each figure shows multi-authorship trends for the mentioned time span, and the right side is for single-authorship trends. As mentioned previously, the lines show collaboration, and an increase in intensity of these lines indicates increase in collaboration. *Thin circles* are representatives of the authors; with varying diameter as an indicator of diversity (or degree). *Thicker circles* are self-loops for single-author trends and clearly the count of such loops is decreasing from 1963-2011; thus, the overall authorship patterns are changing from single- to multi-authorship. Moreover, in earlier decades the authors of single-authored publications had higher degree values where as in the later decades e.g., in 2003-2011 nearly all the nodes of single-authored publication have same smaller diameter. This trend is also apparent from the relative ratio of *single-vertex connected components* to *self-loops* - this ratio is increasing over years.

In table VI, citation counts have been used for defining the prominent authors of both single- and multi-authored articles. It lists the titles, years and authors of the top ten articles during 2003–2011 based on citation counts (as available from ISI Web of Science®). Interestingly, only one single-authored article made to this list. Based on the

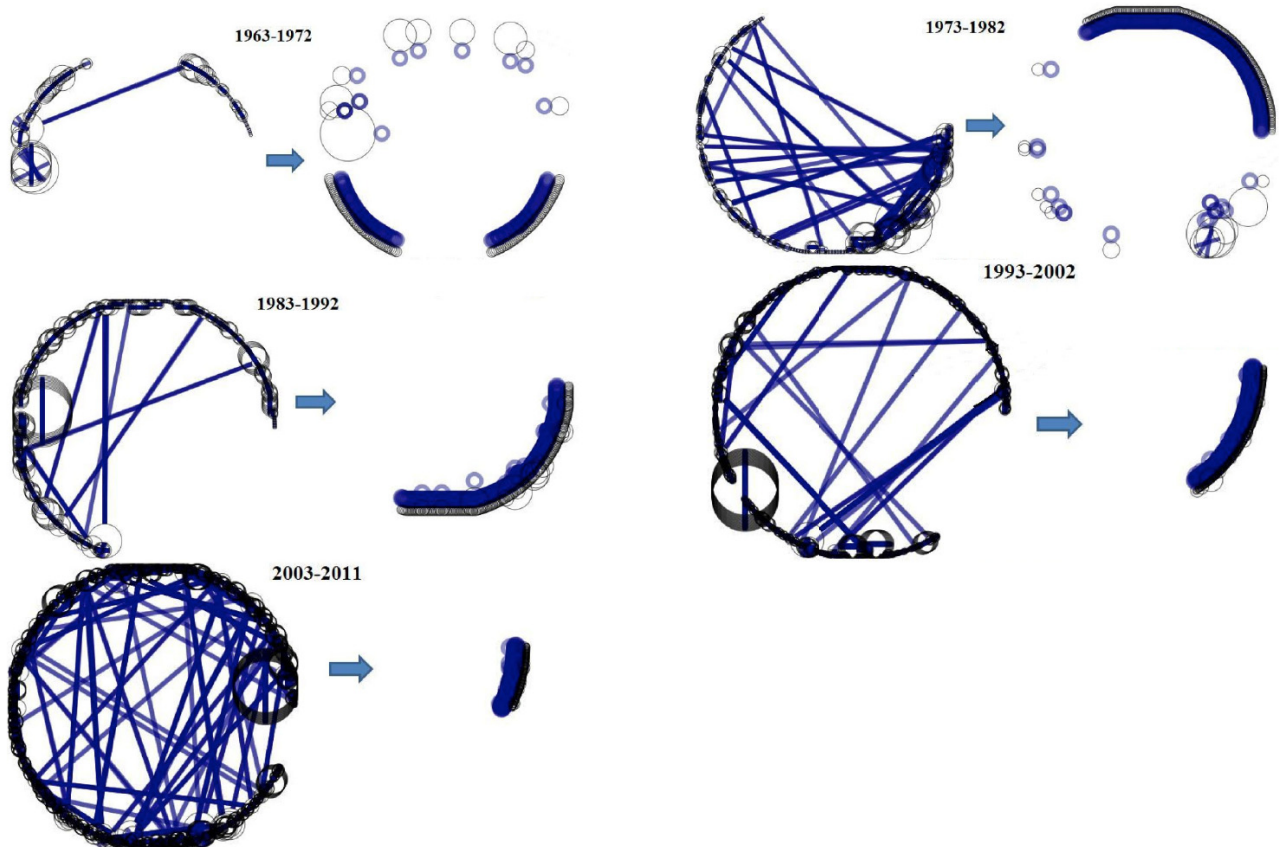


Figure 4. Overall authorship patterns for *IEEE T EDUC*, split into multi-authored (on the left side of each figure) and single-authored articles (on the right side for each decade), 1963-2011

normalized values it is found that the paper by A.C. Graesser *et al.* was cited most frequently and received 7.14 citations per year. Clearly, the articles published in earlier years 2003-2006 have made to the top list due to the fact that they have been available to readers for a longer time. Additionally, key-phrase analysis of the most cited articles was performed to study their research area. It was found that all the articles covered one or more of the following topics: intelligent tutoring system, distance learning, remote laboratory, remote education, web-based education and automatic control etc. Based on this analysis, it can be asserted that these research areas represent the recent trends in EER and online learning can be regarded as one of the primary focus areas. Lastly, these top authors represent a culturally diverse community, their nationalities include: USA, UK, Greece, Italy, Spain, Australia, Ireland, Switzerland etc.⁷

Next, the citation patterns during the last decade of this journal were studied. A count of TC (times cited) is summed up in Table VII. From this table, 68.14% of all the articles (415 out of 609) published during 2003-2011 were cited⁸.

This high percentage is an indicator of the high readership of the journal. Some of the articles from 2011, however, might not be available to the researchers; hence, the citation count for 2011 is expected to rise in the coming years. It is also interesting to see that some articles have been cited more than 50 times while others have been cited only once. Therefore, it was necessary to normalize the

citation count by finding its weighted average, which turns out to be 4.09 citations per article (which is again fairly high).

This high percentage is an indicator of the high readership of the journal. Some of the articles from 2011, however, might not be available to the researchers; hence, the citation count for 2011 is expected to rise in the coming years. It is also interesting to see that some articles have been cited more than 50 times while others have been cited only once. Therefore, it was necessary to normalize the citation count by finding its weighted average, which turns out to be 4.09 citations per article (which is again fairly high).

This completes the citation analysis for the period 2003-2011. The aim of the next section is to relate the already analyzed fields of citation and authorship patterns (collaboration).

TABLE VII.
 CITATION TRENDS FOR *IEEE T EDUC*, 2003-2011

<i>Times Cited</i>	<i>Articles</i>	<i>Times Cited</i>	<i>Articles</i>
0	194	7	19
1	82	8	18
2	68	9	16
3	43	10	10
4	42	11 ≤ TC ≤ 14	30
5	23	15 ≤ TC ≤ 20	18
6	26	20 < TC	20

⁷ Complete author affiliations are provided in supplementary material.

⁸ The data was downloaded in early 2012.

C. Can citation be associated with collaboration?

Table VIII is aimed at seeing if there exist any relationship between citations and collaboration? Since some of the articles were getting cited more than 100 times and some were not getting cited at all, it was necessary to normalize the citation weights in order to compare the two trends. From cited weighted average the overall trend is that: the multi-authored articles get cited more frequently than their single-authored counterparts (even when the total number of single-authored articles is higher, e.g. during 1963–1972, 1973–1982, and 1983–1992).

This phenomenon has also been studied by other researchers [53-54]. Sooryamoorthy in fact states that it is now commonly accepted that co-authorship leads to higher citation rates [55]. One reason for the high citation count could be collaboration and inter-disciplinarity, which means that a multi-authored paper could be the result of collaboration between authors from different disciplines, and hence the article tends to cover a broader scope. Such an article, thus, draws a larger audience and gets cited more often. As T. Bart [56] suggests ‘multi-authorship increases above all the probability to be cited by others’. Furthermore, some of the earlier studies have suggested that research by larger groups tends to be more influential [57-58]. The second reason for high citation counts could be self-citation. In multi-authored articles, self-citation occurs when at least one author is common between the citing and the cited articles [59]. Thus, the more the number of authors per article, the higher is the probability of getting self-citations [56]. This completes the analysis regarding main contributors, authorship trends, collaboration, and citation. The next section covers questions regarding the themes within *IEEE T EDUC*.

IV. CONTENT ANALYSIS

A. What are the past themes that have been explored in this journal, and what are the new emerging themes?

This question will be answered by studying the trends that have been taking place in this journal since its inception in 1963. For this purpose the keyword (*DE and ID*) field [and abstracts (*AB*) for key-phrase and topic modeling], which indicates ‘core concepts and central fields of concern’ [60] will be used. In the absence of the keyword field, it was decided to use the *titles (TI field)* for keyword extraction due to correlation values calculated⁹. The relevance of titles as a source of keywords has been discussed in [61-67]. Like the authorship analysis, a more comprehensive study is presented for the last time span i.e., in addition to the keyword and key-phrase analysis, word co-occurrence and topic modeling techniques will be applied to get better context of research topics.

1) Content Analysis:

Content analysis is a systematic and replicable approach for reducing many pieces of data into more relevant and manageable contents based on explicit rules of coding [68-72]. Holsti [73] defines content analysis in a broader way as “any technique for making inferences by objectively and systematically identifying specified characteristics of messages.” According to Palmquist [74] “it is used to determine the presence of certain words, concepts, themes, phrases, characters, or sentences within texts or sets of

⁹ The methodology for finding the correlation values is presented in the supplementary material of this paper.

TABLE VIII.
CITATION - COLLABORATION TRENDS FOR *IEEE T EDUC*, 1963-2011

Year	Authorship	Total	Not Cited Average	Cited Average	Cited Weighted Average
1963–1972	single-author	224	71.43	28.57	0.558
	multi-author	96	63.54	36.46	1.771
1973–1982	single-author	201	55.22	44.78	1.189
	multi-author	139	43.88	56.12	1.719
1983–1992	single-author	203	26.6	73.4	3.31
	multi-author	182	25.82	74.18	3.7
1993–2002	single-author	217	28.11	71.89	3.456
	multi-author	378	15.34	84.66	6.272
2003–2011	single-author	129	33.33	66.66	3.14
	multi-author	480	31.46	68.54	4.34

texts and to quantify this presence in an objective manner”. The aim of content analysis in this article is to identify the latest research trends therefore, word co-occurrence analysis and topic modeling is performed on the abstracts of the articles during the last decade of the journal.

2) Frequency Analysis:

Frequency analysis – the concept of using keyword counts, has been used for similar studies by many researchers. The question is: How well can this method predict the research trends? As Summers [75] suggests, “all aspects of lexicography are influenced by frequency.” Similarly, Kilgarriff [76] is of the opinion that the significance of a word is known by its usage. The more common it is, the more important it is to know. R.A. Bentley [77] suggests: “The evolution of vocabulary in academic publishing is characterized via keyword frequencies recorded in the ISI Web of Science citations database.” Other authors have also suggested the importance of frequency-based analysis for the more frequently occurring words [79-80], so usage of frequency-based analysis is justified. In this paper the usage of frequency based analysis is for studying the evolution of EER as a field and for this purpose keyword and key-phrase analysis is performed. A similar study for the field of Marketing Science which relies on the usage of keyword counts can be found at [81]. A. Duvvuru *et al.* [82] conducted a keyword based study for the European Journal of Operational Research. They suggested a network analysis technique for the keywords in scholarly articles. X. Wang *et al.* [83] conducted a study for the Journal of *Scientometrics* where they proposed a real time method for detecting the emerging research trends by monitoring the downloads of research articles of this journal and aggregating it with the article keywords.

3) Methodology:

This methodology relies on the usage of keywords and key-phrases and attempts to see their spread in the five decades. Regarding, the keyword rank and classification as presented in this study, it is suggested that: “academic keywords should be ranked in order of popularity from year to years...to see what’s hot and what’s not [77]” To generalize the results of this study across the EER field, comparisons are made to earlier studies [7-9, 78] in engineering education that utilized keywords for predicting the trends of this community. As mentioned earlier, all the analysis is done only on the keywords field (DE, ID fields from ISI WoS) and when this field is absent the keywords

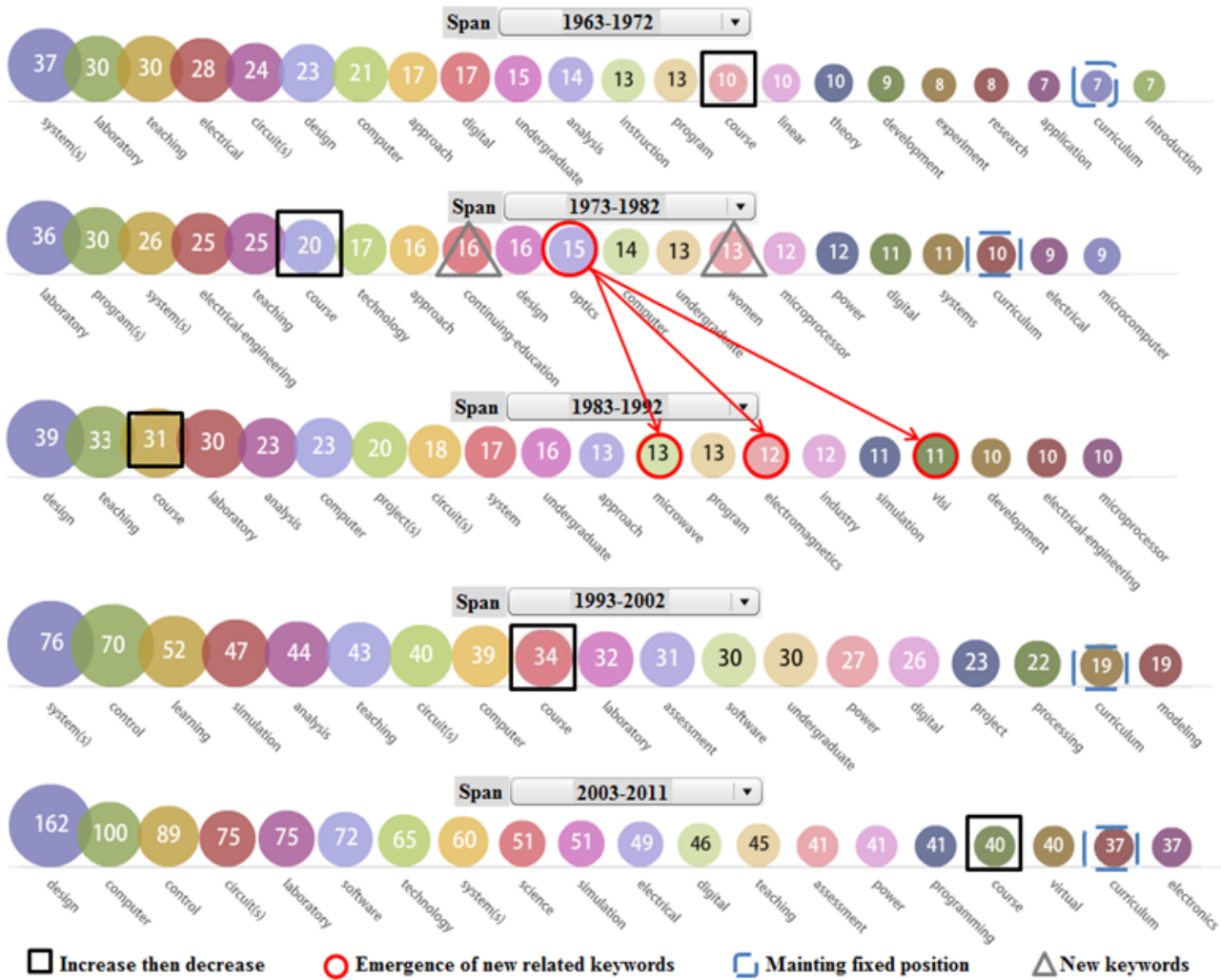


Figure 5. Overall research theme shifts in *IEEE T EDUC*, 1963-2011 (split into decades)

are extracted from the titles (TI field of ISI WoS). Once the titles and keywords have been parsed out of the articles, frequency analysis is then used for the purpose of extracting top keyword during each decade by using the Hermetic Word Frequency Counter (HWFC) software. HWFC first processes the data to eliminate some 353 common English words and then generates the top keywords. After this, following refinements were made in the top keywords list: elimination of obvious keywords with reference to the scope of the journal (like *engineer*, *education*, *students*, etc.), elimination of too broad or general terms (such as *new*, *paper*, *presents*, etc.), and merging of word variants (*learn[ing]*, *lab[s]*, *laboratory*, *design[s]*, *designed*, etc.). Once the top 20 keywords during each decade were obtained, they were plotted using the IBM software tool Many Eyes, as shown in Figure 5. The size of each keyword is based on the number of times it occurred. This number is also written inside each circle – however, rather than absolute counts, the relative rank or position of the keywords in the given time span will be used to answer questions related to trends. Some of the keywords maintain their overall rank during each decade, such as *curriculum* (as shown by dashed blue squares). During the third decade, it appears that the usage of *curriculum* might have decreased since it did not make to the top keyword list. But it was found that this word was still used as many times as eight (8). So it can be said that the

topics pertaining to curriculum development have consistently retained their rank within this journal. The keyword *curriculum* also made to the top list for the *JEE* in [7-8] between the years 1993-2002 (in these studies the articles of the *JEE* were manually coded for classification). Using the key-phrase analysis, it was found that *curriculum* has been used in the context of engineering curriculum, undergraduate curriculum, curriculum planning, design and development, Elect. Engg and ECE curriculum etc.

Similarly, it can also be concluded that topics pertaining to diversity, such as women and minority education, as well as continuing education, gained more interest during the second decade of this journal, as evident from the keywords (shown by grey triangles). Interestingly, these topics did not make to the top list of *IJEEE* [9], however, for *JEE* these topics made to the top list [7-8]. Key-phrase analysis of *IEEE* revealed that these keywords have been used for gender differences, gender race retention and women in engineering etc.

Keywords highlighted by red circles show the emergence of new themes, such as when *optics* appeared in the top 20 list during the second decade and when, in the third decade, some related keywords such as *VLSI*, *microwave*, *electromagnetics*, etc., appeared in the top list. Key-phrase analysis revealed that *optics* has been used in the context of *fiber optics*, *photonics* and *lasers* etc. Interestingly

PAPER
AUTHORSHIP AND CONTENT ANALYSIS OF ENGINEERING EDUCATION RESEARCH: A CASE STUDY

TABLE IX.
TOP TEN TOPICS FOR *IEEE T EDUC*, 2003-2011

Topic	Top ten words in the topics
1	control power simulation presented implementation matlab modeling electric designed undergraduate
2	study results academic technology effective higher instructional groups impact strategies
3	learn teachers active traditional classroom collaborative distance satisfaction online motivation
4	problem model proposed data practice signals measured space designing characteristics
5	student assessment test group improve strategy ee online concept quality
6	students experience skills information learn communication technologies application knowledge mobile
7	modern measurement fundamental technique topic understanding mathematical algorithms theoretical equations
8	web educational software environment tool open interactive principles oriented authors
9	programming teaching designed knowledge modules language materials learned game framework
10	design digital hardware field processing signal logic techniques complex integrated
11	engineering curriculum education school technology university science faculty ece outreach
12	computer courses teaching architecture science years assignments pbl instruction organization
13	laboratory experiments virtual internet remote network platform access lab wireless
14	results evaluation methodology education pedagogical introduction result goal found approaches
15	program research undergraduate students computing topics material graduate wide curricula

V. CONCLUSIONS AND DISCUSSION:

The authors would like to start this section with a comment that this study is data-driven (scientometric study). Here the structure and content of the publications of *IEEE T Educ* served as a guideline for the analysis. Such data driven methods have been used in other disciplines too such as in Human-Computer Interaction (HCI) by N. Henry [85]. He suggests that peer reviewed publications are one of the most fundamental tools for communicating and accessing results in any learning and scientific society. Therefore, this study has utilized the publications (and associated citation data) of *IEEE* to make known the evolution and patterns of this journal and the field of EER. As S. Redner [86] suggests, citation data can be used to identify influential research and researchers, new potential in research, unforeseen connections across fields, and downturns of the subfields that are exhausted.

In this study only citation counts were considered and the citing or cited sources were not studied. For *IEEE* these sources are already studied by Wankat [11] and Williams [12]. These, and other such studies [87-89] revealed that there is limited cross fertilization of engineering education research and development (R&D). There have also been some other studies that highlighted the slow rate of dissemination of EER [89-90]. Therefore, the content analysis as presented in this paper may help dissemination of EER by highlighting the emerging trends in this area for interested engineering faculty to follow. Furthermore, due to the interdisciplinary nature of the field of EER the authors believe that this study can benefit not only the members of this field but also those who interact with them. Rookies in EER may find a road map to its milestone research, main trends and central authors. Expert researchers in EER can get a global overview to help them gain better insight about their own role and their colleague's role in shaping this field. Lastly, for researchers outside the field of EER (such as in the field of information systems and scientometrics etc.), such study can provide context for comparing the field of EER with other disciplines. Based on the analysis performed in this paper, the followings can be concluded: (1) nearly 55% of the contributions are from USA (2) although there have been

only 6% contributions from Spain and yet it dominates the top list of authors (3) analysis of authorship trends showed the growth of this journal in terms of publication counts and in terms of distinct authors who contribute to the journal. The reason for an increase in author counts was found to be collaboration as evident from the fact that now multi-author trends are dominant. It is evident that the general trend for authorship is shifting towards collaboration and it is a general trend followed these days [50-52] (4) within this journal community is shaping as close-knit, showing that (due to interdisciplinary nature of EER) collaboration is increasing (5) although the journal seems to be shaping as well connected and dense community with increasing collaboration trends, however, it was found that co-authorship between same author pairs is not very common (6) this study also revealed that the multi-authored articles get cited more frequently than single-authored publications possibly due to self-citations and/or the interdisciplinary nature of multi-authored articles (7) topic modeling techniques revealed that *online learning*, *assessment*, *design* and *modeling* are the among the dominant study areas within *IEEE*, as well as EER [78] (8) lastly, based on the in-depth analysis of contents it was found that the themes found in *JEE*, *FIE* (Frontiers in Educ.), *IJEE* (Int. Journal of Engg. Educ.) and *IEEE* all indicate similar results and thus study of such communities may help researchers in understanding the evolution, growth and future of the EER discipline.

VI. LIMITATIONS:

Like other studies, this study has some caveats. Firstly, the current trends in the field of citation analysis are due to the data availability from Science Citation Index (SCI) by the ISI [92]. One limitation of SCI is although it is the biggest database but it does not cover all the scientific and technical venues; nor it intends to do so [93]. Secondly, 5-10% of citations can be flawed (due to spelling errors etc.) [94-95]. Third, high citation of an article could be due to various reasons while some are practical (such as: to introduce a study that builds on previous knowledge, to refer to methodologies already developed or utilized in previous articles [96]) but others are doubtful (such as excessive self-citation and 'citation of poor work' [97] etc.). So

citation count can only provide ‘an approximate measure for scientific quality’ [85]. A visual limitation of this analysis is based on the selected software tool NodeXL, in which the single-author patterns are represented by self-loops. However, this limitation is just visual and has no effect on the data and results. In some of the earlier studies the nature of citing/cited sources for EER venues have been studied. However, due to broad range of topics covered in this paper, this analysis is restricted only to citation counts. Lastly, in this paper the bibliometric data has been taken from a single database (WoS).

FUTURE WORK AND IMPLICATIONS

To extend the current work, interested reader may also explore questions like the followings: What are the factors controlling the co-authorship network? Can research be associated with authorship/co-authorship?

REFERENCES

- [1] N.L. Osorio and M.A. Osorio, "Engineering education in Europe and the USA," *Science & Technology Libraries*, 23:1, 49-70, 2005. http://dx.doi.org/10.1300/J122v23n01_05
- [2] N.L. Osorio, "What every engineer should know about engineering education," in proceedings of American Society of Engineering Education (ASEE), DeKalb, Illinois, 2005.
- [3] A.L. Porter and J. Youtie, "How interdisciplinary is nanotechnology?" *J. Nanopart. Res.*, Springer, 2009. <http://dx.doi.org/10.1007/s11051-009-9607-0>
- [4] Denton, Denise D. "Engineering Education for the 21st Century: Challenges and Opportunities." *Journal of Engineering Education*, 87(1): 19-22, 1998. <http://dx.doi.org/10.1002/j.2168-9830.1998.tb00317.x>
- [5] NEERC (National Engineering Education Research Colloquies) Special Report "The Research Agenda for the New Discipline of Engineering Education," *Journal of Engineering Education*, 94(4), pp. 257-261, 2006
- [6] R. Streveler and K. Smith, Guest Editorial: From the Margins to the Mainstream: "The Emerging Landscape of Engineering Education Research," *J. Eng. Educ.*, vol. 99(4), pp. 285-287, 2010 <http://dx.doi.org/10.1002/j.2168-9830.2010.tb01063.x>
- [7] P.C. Wankat, "Analysis of the First Ten Years of the Journal of Engineering Education," *Journal of Engineering Education*, 93(1), 2004
- [8] P.C. Wankat, "An Analysis of the Articles in the *Journal of Engineering Education*," *Journal of Engineering Education*, pp. 37-42, 1999
- [9] S. Nawaz *et al.* "Analysis of the influence of the International Journal of Electrical Engineering Education on electrical engineering and electrical engineering education" *International Journal of Electrical Engineering Education*, vol. 50, no. 3, pp 312-336, 2013.
- [10] J. Strobel *et al.*, "Is the Engineering Education Community Becoming More Interdisciplinary?" In Proceedings of American Society of Engineering Education (ASEE) Annual Conference & Exposition, San Antonio, TX, 2012.
- [11] P.C. Wankat, "Guest Editorial: Cross-Fertilization of Engineering Education Research and Development" *IEEE TRANSACTIONS ON EDUCATION*, vol. 54, no. 4, 2011. <http://dx.doi.org/10.1109/TE.2011.2165757>
- [12] B. Williams and P. Neto, "Tracking engineering education research and development – contributions from bibliometric analysis," *International Journal of Engineering Pedagogy*, vol. 2, no. 2, 2012.
- [13] B. Williams and M. Alias, "Strategic Pathways to Engineering Education Research: case study of a top-down initiative", in *Proc. of the Research in Engineering Education Symp. (REES 2011)*, Madrid, Spain, October, 2011.
- [14] I. X. D. Library. IEEE Transactions on Education [Online]. Available: I. X. D. Library. IEEE Transactions on Education [Online]. Available: <http://ieeexplore.ieee.org/xpl/RecentIssue.jsp?punumber=13>
- [15] B. K. Jesiek, M. Borrego, K. Beddoes, M. Hurtado, P. Rajendran and D. Sangam, "Mapping Global Trends in Engineering Education Research, 2005–2008," *Int. J. Eng. Educ.*, vol. 27, n. 1., pp. 77-90, 2011.
- [16] E. De Graaff and A. Kolmos, "Research Methods In Engineering Education Research," in *Proc. of the Joint International IGIP-SEFI Ann. Conf. 2010*, Trnava, Slovakia, 19th - 22nd Sept. 2010.
- [17] L.I. Meho and K. Yang, "Impact of Data Sources on Citation Counts and Rankings of LIS Faculty: Web of Science versus Scopus and Google Scholar," *Journal of the American Society for Information Science and Technology*, 58(13):2105–2125, 2007. <http://dx.doi.org/10.1002/asi.20677>
- [18] K.M. Whitley, "Analysis of SciFinder Scholar and Web of Science citation searches," *Journal of the American Society for Information Science and Technology*, 53(14), 1210–1215, 2002. <http://dx.doi.org/10.1002/asi.10192>
- [19] A.W. Jones, "The distribution of forensic journals, reflections on authorship practices, peer review and role of impact factor," *Forensic Science International*, Elsevier, 2007.
- [20] L. Borrmann *et al.* "Convergent validity of bibliometric Google Scholar data in the field of chemistry—Citation counts for papers that were accepted by *Angewandte Chemie International Edition* or rejected but published elsewhere, using Google Scholar, Science Citation Index, Scopus, and Chemical Abstracts," *Journal of Informetrics*, vol. 3, pp. 27–35, 2009.
- [21] P. Jacsó, "As we may search—comparison of major features of the Web of Science, Scopus, and Google Scholar citation-based and citation enhanced databases," *Current Science*, 89(9), 1537–1547, 2005.
- [22] K. Bauer and N. Bakkalbasi, "An examination of citation counts in a new scholarly communication environment," *D-Lib Magazine*, 11(9), 2005. <http://dx.doi.org/10.1045/september2005-bauer>
- [23] H.F. Moed, "New developments in the use of citation analysis in research evaluation," *Archivum immunologiae et therapeutiae experimentalis*, Springer, 2009.
- [24] H.F. Moed, *Citation analysis in research evaluation*, Kluwer Academic Publishers, Dordrecht, 2005.
- [25] C. Neuhaus and H.D. Daniel, "Data sources for performing citation analysis—an overview," *Journal of Documentation*, 64(2), 193–210, 2008 <http://dx.doi.org/10.1108/00220410810858010>
- [26] Last accessed: Oct 5, 2013. Available: <http://thomsonreuters.com/web-of-science/>
- [27] J.E. Froyd, P.C. Wankat and K.A. Smith, "Five major shifts in 100 years of engineering education," *Proceedings of the IEEE*, 2012 <http://dx.doi.org/10.1109/jproc.2012.2190167>
- [28] N.D. Bellis, *Bibliometrics and citation analysis: from the Science Citation Index to cybermetrics*. Lanham, MD: Scarecrow Press Inc., 2009.
- [29] R. Klavans and K. W. Boyack, "Thought leadership: a new indicator for national and institutional comparison" *Scientometrics*, vol. 75, pp. 239–250, 2008. <http://dx.doi.org/10.1007/s11192-007-1854-1>
- [30] T. Heinze and G. Bauer, "Characterizing creative scientists in nano-S&T: productivity, multidisciplinary, and network brokerage in a longitudinal perspective" *Scientometrics*, vol. 70, pp. 811–830, 2007. <http://dx.doi.org/10.1007/s11192-007-0313-3>
- [31] D. K. Simonton, *Origins of genius: Darwinian perspectives on creativity* New York: Oxford University Press, 1999.
- [32] D. K. Simonton, *Creativity in science: chance, logic, genius, and zeitgeist* Cambridge: Cambridge University Press, 2004. <http://dx.doi.org/10.1017/CBO9781139165358>
- [33] R. Pei and A. L. Porter, "Profiling leading scientists in nanobiomedical science: interdisciplinary and potential leading indicators of research directions," *R&D Management*, vol. 41, pp. 288-306, 2011. <http://dx.doi.org/10.1111/j.1467-9310.2011.00643.x>
- [34] T. Heinze, *et al.*, "Identifying creative research accomplishments: methodology and results for nanotechnology and human genetics," *Scientometrics*, vol. 70, pp. 125–152, 2007. <http://dx.doi.org/10.1007/s11192-007-0108-6>
- [35] Last accessed: Oct 6, 2013. Available: <http://www8.cs.umu.se/kurser/TDBAfl/VT06/algorithms/BOOK/BOOK/NODE4.HTM>

- [36] J. Greiner. (2012). Graph Connectivity. Last accessed: February 20, 2012. Available: <http://www.cs.cmu.edu/~scandal/alg/connectivity.html>
- [37] D. Jungnickel, *Graphs, networks and algorithms*, vol. 5, Springer, 2008. <http://dx.doi.org/10.1007/978-3-540-72780-4>
- [38] D. West, *Introduction to Graph Theory*, (2nd Edition), Prentice Hall, 2000.
- [39] M. Smith, N. Milic-Frayling, B. Shneiderman, E. Mendes Rodrigues, J. Leskovec, C. Dunne, "NodeXL: a free and open network overview, discovery and exploration add-in for Excel 2007/2010," <http://nodexl.codeplex.com/> from the Social Media Research Foundation, <http://www.smrfoundation.org>, 2010
- [40] OXFORD DICTIONARIES, 2013. Oxford dictionary collaboration. Last accessed: 27th March, 2013. Available: <http://oxforddictionaries.com/definition/english/collaboration>
- [41] M. Smith, 1958, The trend toward multiple authorship in psychology, *American Psychologist* 13, 596-599. <http://dx.doi.org/10.1037/h0040487>
- [42] J.S. Katz and B.R. Martin, "What is research collaboration?" *Research Policy* 26, pp. 1-18, Elsevier Science B.V. 1997
- [43] K. Subramanyam, "Bibliometric studies of research collaboration: A review," *Journal of Information Science* 6, 35, 1983 <http://dx.doi.org/10.1177/016555158300600105>
- [44] C. Balog, "Multiple authorship and author collaboration in agricultural research publications," *Journal of Research Communication Studies* 2, 159-169, 1979/80
- [45] D.de B. Beaver and R. Rosen, "Studies in scientific collaboration: Part I--The professional origins of scientific co-authorship," *Scientometrics* 1, 65-84, 1978. <http://dx.doi.org/10.1007/BF02016840>
- [46] D.de B. Beaver and R. Rosen, "Studies in scientific collaboration: Part II--Scientific co-authorship, research productivity and visibility in the French scientific elite, 1799- 1830," *Scientometrics* 1, 133-149, 1979. <http://dx.doi.org/10.1007/BF02016966>
- [47] D.de B. Beaver and R. Rosen, "Studies in scientific collaboration: Part III--Professionalization and the natural history of modern scientific co-authorship," *Scientometrics* 1, 231-245, 1979. <http://dx.doi.org/10.1007/BF02016308>
- [48] B.L. Clarke, "Multiple authorship trends in scientific papers," *Science* 143, 822-824, 1964 <http://dx.doi.org/10.1126/science.143.3608.822>
- [49] A.J. Meadows and J.G. O'Connor, "Bibliographic statistics as a guide to growth points in science," *Science Studies* 1, 95-99, 1971
- [50] M. Greene, "The demise of the lone author," date accessed: Nov. 2013 <http://dx.doi.org/10.1038/nature06243>
- [51] C. King, "Multi-author Papers: Onward and Upward," Science-Watch Newsletter, July 2012. Accession date: Nov. 2013, http://archive.sciencewatch.com/newsletter/2012/201207/multi-author_papers/
- [52] A. Regalado, "Multi-author papers on the rise," *Science*, <http://dx.doi.org/10.1126/science.7701334>
- [53] S.M. Lawani, "Some bibliometric correlates of quality in scientific research," *Scientometrics* 9, 13-25, 1986. <http://dx.doi.org/10.1007/BF02016604>
- [54] L.A. Hurley *et al.* "Deconstructing the collaborative impact: Article and author characteristics that influence citation count"
- [55] R. Sooryamoorthy, "Do Types of Collaboration Change Citation? Collaboration and Citation Patterns of South African Science Publications," *Scientometrics* 81.1, 177-93 (2009). <http://dx.doi.org/10.1007/s11192-009-2126-z>
- [56] T. Bart and W. Glänzel, "Does co-authorship inflate the share of self-citations?" *Scientometrics*, vol. 61, no. 3, pp. 395-404, 2004.
- [57] D. Crane, "Invisible Colleges" (University of Chicago Press) , 1972.
- [58] W. Goffman and K.S. Warren, "Scientific Information Systems and the Principle of Selectivity" (Praeger, New York), p. 127, 1980.
- [59] H. Snyder, and S. Bonzi, "Patterns of self-citation across disciplines," *Journal of Information Sciences*, 24 (1998), 431-435. <http://dx.doi.org/10.1177/016555159802400606>
- [60] B. Dutta *et al.* "Classification of keywords extracted from search articles published in science journals," *Annals of Library and Information Studies*, vol. 55, pp. 317-333, 2008.
- [61] R.T. Bottle, "The information content of titles in engineering literature," *IEEE Transactions on Engineering Writing*, EWS-13 (1970) 41-45. <http://dx.doi.org/10.1109/TEWS.1970.4322437>
- [62] I. Hansen, "Evaluation of the databases CA condensates compared with chemical titles," *Journal of Chemical Documentation*, 12 (1972) 101-109. <http://dx.doi.org/10.1021/c160045a012>
- [63] D.H. Kraft, "Comparison of keyword-in-context indexing of titles with a subject-heading classification system," *American Documentation*, 15 (1964) 48-52. <http://dx.doi.org/10.1002/asi.5090150109>
- [64] F.W. Lancaster, *Vocabulary control for information retrieval*, (Information Resources Press; ashington DC), 1972.
- [65] G. Olive *et al.* "Studies to compare retrieval using titles with that using index terms," *Journal of Documentation*, 29 (1973) 169-191. <http://dx.doi.org/10.1108/eb026554>
- [66] M.J. Ruhl, "Chemical documents and their titles: Human concept indexing vs. KWIC machine indexing," *American Documentation*, 15 (1964) 136-141. <http://dx.doi.org/10.1002/asi.5090150213>
- [67] J. Kaur and V. Gupta, "Effective approaches for extraction of keywords," *International Journal of Computer Science Issues*, vol. 7, no. 6, 2010.
- [68] B. Berelson, *Content Analysis in Communication Research*. . Glencoe, Illinois: Free Press, 1952.
- [69] S. Stemler. (2001, 23rd March, 2012). An Introduction to Content Analysis. *ERIC Digest*. Available: <http://www.ericdigests.org/2002-2/content.htm>
- [70] K. Krippendorff, *Content Analysis: An Introduction to Its Methodology*. Newbury Park, CA: Sage, 1980.
- [71] U. S. G. A. Office, *Content Analysis: A Methodology for Structuring and Analyzing Written Material*. . Washington, D.C., 1996.
- [72] R. P. Weber, *Basic Content Analysis*, 2nd ed. Newbury Park, CA, 1990.
- [73] O. R. Holsti, *Content Analysis for the Social Sciences and Humanities*. . Reading, MA: Addison-Wesley, 1969.
- [74] M. Palmquist. (25th September, 2013). Content Analysis. Available: <http://www.colostate.edu/Depts/WritingCenter/references/research/content/page2.htm>
- [75] D. Summers, " Computer lexicography: the importance of representativeness in relation to frequency.," in *Using Corpora for Language Research*, J. Thomas and M. Short, Eds., ed. London: Longman, 1996.
- [76] A. Kilgariff, "'Putting Frequencies in the Dictionary'. ," *International Journal of Lexicography* vol. 10, pp. 135-155, 1997. <http://dx.doi.org/10.1093/ijl/10.2.135>
- [77] R.A. Bentley, "Random Drift versus Selection in Academic Vocabulary: An Evolutionary Analysis of Published Keywords," *PLoS One*, vol. 3, no. 8, e3057, 2008. <http://dx.doi.org/10.1371/journal.pone.0003057>
- [78] K. Madhavan *et al.*, "Understanding the Engineering Education Research Space Using Interactive Knowledge Networks and Topic Modeling Techniques," IKNEER Report, 2011.
- [79] A. Kilgariff and R. Salkie, "'Corpus Similarity and Homogeneity via Word Frequency'. ," in *EURALEX '96*, Gothenberg, Sweden, 1996, pp. 121-130.
- [80] T. J. Otlogetswe, "Using frequency and keyword analysis to extract candidates for setswana dictionary labels," *International Journal of Lexicography*, 2012. <http://dx.doi.org/10.1093/ijl/ecs001>
- [81] C.F. Mela *et al.* "A keyword history of Marketing Science," November 5, 2012.
- [82] A. Duvvuru *et al.*, "Undercovering research trends: Network analysis of keywords in scholarly articles," Ninth International Joint Conference on Computer Science and Software Engineering (JCSSSE), 2012. <http://dx.doi.org/10.1109/JCSSSE.2012.6261963>
- [83] X. Wang *et al.*, "Tracing scientist's research trends realtimely," *Scientometrics*, 2012. <http://dx.doi.org/10.1007/s11192-012-0884-5>
- [84] Sci2 Team. (2009, Science of Science (Sci2) Tool. Available: <http://sci2.cns.iu.edu>
- [85] N. Henry *et al.* "20 years of four HCI conferences: A visual exploration," *International Journal of Human-Computer Interaction*, vol. 23, no. 3, pp. 239-285

- [86] S. Redner, "Citation Statistics from 110 Years of Physical Review," *Physics Today*, vol. 58, pp. 49-54, 2005. <http://dx.doi.org/10.1063/1.1996475>
- [87] P. C. Wankat, "Guest editorial: Cross-fertilization of STEM education communities," *J. STEM Educ.*, 2011.
- [88] P. C. Wankat, "Guest editorial: Cross-fertilization of engineering education R & D," *Chem. Eng. Educ.*, vol. 45, no. 4, p. 230, 2011.
- [89] P. C. Wankat, "Forum: Cross-fertilization of engineering education research and development," *J. Prof. Issues Eng. Educ. Practice*, 2011.
- [90] M. Borrego, J.E. Froyd and T.S. Hall, "Diffusion of engineering education innovations: A survey of awareness and adoption rates in US engineering departments," *J. Eng. Educ.*, 99(3), pp. 185-207, 2010. <http://dx.doi.org/10.1002/j.2168-9830.2010.tb01056.x>
- [91] P.C. Wankat, "Cross fertilization of engineering education research and development," *Journal of professional issues in engineering education & practice*, ASCE, 2012.
- [92] G. Pinski and F. Narin, "Citation influence for journal aggregates of scientific publications: Theory, with application to the literature of physics," *Information Processing & Management*, vol. 12 pp. 297-312, 1976. [http://dx.doi.org/10.1016/0306-4573\(76\)90048-0](http://dx.doi.org/10.1016/0306-4573(76)90048-0)
- [93] E. Garfield, "Citation Analysis as a Tool in Journal Evaluation," *Science*, vol. 178, pp. 471-479, 1972. <http://dx.doi.org/10.1126/science.178.4060.471>
- [94] H. F. Moed, "Towards a critical, informative, accurate and policy relevant bibliometrics," *Nature*, vol. 415, pp. 731-732, 2002. <http://dx.doi.org/10.1038/415731a>
- [95] S. Redner. Last accessed: January 22, 2011. Available: <http://arxiv.org/abs/physics/0407137v1>
- [96] W. Glänzel *et al.*, "A concise review on the role of author self-citations in information science, bibliometrics and science policy," *Scientometrics*, 67(2), 263-277, 2006. <http://dx.doi.org/10.1007/s11192-006-0098-9>
- [97] A. Smith and M. Aysenck, "The correlation between ratings and citation counts in psychology," 2002.
- [98] Andrew K. McCallum "MALLET: A Machine Learning for Language Toolkit." <http://mallet.cs.umass.edu>. 2002

AUTHORS

Sadia Nawaz is PhD student in the school of Computing and Information Systems at the University of Melbourne (UniMelb), Australia. Prior to starting her PhD, she worked as RA in the Learning Environments of UniMelb. She holds Master of Science in Electrical and Computer Engineering from Purdue University, Indiana, USA (e-mail: sadia.nawaz@unimelb.edu.au).

Johannes Strobel is with University of Missouri, Columbia, MO, USA. He is Professor in the School of Information Science & Learning Technologies (iSchool) (e-mail: strobelj@missouri.edu).

This work is supported by National Science Foundation (NSF) funding under the grant number 0943198. Any opinions, findings, conclusions, or recommendations expressed in this material are those of the authors and do not reflect the views of the NSF. Submitted, 24th Feb 2016. Published as resubmitted by the authors on 7th April 2016.