

A Scientometric Methodology Based on Co-Word Analysis in Gas Turbine Maintenance

Ali NEKOONAM, Reza Fatehi Nasab, Soheil JAFARI, Theoklis NIKOLAIDIS, Nader ALE EBRAHIM, Seyed Alireza MIRAN FASHANDI*

Abstract: Evaluation of scientific journals has a profound effect on the future of scientific research so that different institutes and countries can set appropriate goals and invest with less risk in various scientific fields. Accordingly, this article presents a new method based on a combination of co-word analysis and social network analysis to extract the hotspot topics. Using HistCite, NodeXL, and VOSviewer, then combining their results, the desired analysis is conducted for six time periods. Based on the bibliographic parameters in HistCite and by defining an index, the first five periods are selected such that both quantity and quality of articles in each period are maximum compared to other years, while the sixth time period contains the latest research. For each of the six periods, the co-word networks as created in VOSviewer are analyzed. Next, based on a combination of network centralities developed in NodeXL, the hotspot keywords are specified which are then validated and aggregated using the bibliographic parameters in HistCite. The results reveal five important time periods in gas turbine maintenance. The hotspot keywords obtained for the last period show that in recent years, some topics including gas turbine fault prognosis, neural network-based approaches, big data analysis, sensor fault diagnosis, blade availability, economic analysis and useful life estimation are prominent subjects in gas turbine maintenance.

Keywords: co-word analysis; gas turbine maintenance; HistCite; NodeXL; social network analysis; VOSviewer

1 INTRODUCTION

Scientometrics is a relatively novel branch of science that seeks to measure and analyze scientific products in a bid to simplify the planning strategies of organizations, provide effective management of financial and human resources, increase the efficacy of scientific institutes and guide them in the right direction. This field of science dates back to the 1970s [1] and has been widely used for the analysis and evaluation of scientific literature [2-8].

To date, scientometrics methods have been applied to different research problems, but not yet to gas turbine maintenance. This field of research has a pivotal role in the mechanical and aerospace sectors as it deals with cost, time saving and risk management in powerplants, marine applications, and air transportation. Gas turbine maintenance covers a wide range of various aspects of gas turbine operation including gas turbine monitoring [9], fault detection and identification systems [10], prognostics [11], and system design and control [12-16]. Therefore, scientometrics methods could be applied to manifest the important future research challenges, determine main approaches, and represent an evolutionary trend of the technology over time.

In this study, through a new methodology, co-word network analysis and SNA are combined so as to present hotspot topics and evolutionary trends in the field of gas turbine maintenance from January 1985 to December 2019 over different periods. This methodology relies on the combined capabilities of three research tools, namely NodeXL, VOSviewer, and HistCite and uses two indexes. The considered years are divided into six important periods by applying the first index which is defined based on bibliographic parameters that were extracted from HistCite software. Moreover, a separate analysis is performed for each period and the top hot keywords of each one are presented by introducing the second index which is defined based on combination of network centralities. To this aim, after collecting the data from Web of Science, they are first imported into VOSviewer and the networks are created based on co-occurrence data. Next, the network data is exported to NodeXL in order to obtain network centrality

indexes. In this article, 4 network centralities are considered including betweenness centrality, degree centrality, closeness centrality and eigenvector centrality. Finally, by combining the centralities and defining an overall index, the prominent and hot keywords of the network in each period are obtained. In parallel, the analyses performed on the data in HistCite are exploited for the purpose of results validation and integration.

The structure of the article is as follows. Section 2 provides an overview on the research in the field of co-word analysis with a focus on the publications of recent years. Section 3 introduces the methodology, the data required for analysis, the definition of network centralities and indexes and the employed tools while the investigated time periods are identified. Section 4 presents the results of the analyses. Finally, section 5 discusses the article's conclusions.

2 LITERATURE OF CO-WORD ANALYSIS

Co-word analysis was first introduced by Callon in 1983. The main assumption of this technique is that the keywords of an article provide an acceptable description of its content [17]. Co-word analysis determines the relationship between ideas and concepts [18]. The appearance of two words in a scientific text indicates the existence of a relationship between the topics to which these two words allude. This method has been exploited in the scientometric studies of different fields and for many subjects including environmental footprint, green buildings [19], information behaviour [20], international management science [21], market research [22], environmental acidification [23], psychiatry [24] and life cycle assessment [25].

In recent years, certain methods and new approaches have been proposed by researchers for the co-word analysis [26]. Zhang et al. used co-word analysis alongside multi-dimensional scaling analysis and SNA to determine frequently-studied topics and research opportunities and to depict the evolutionary trend of the creativity domain [27]. Instead of using the conventional technique of word repetition, Wang et al. developed a new method based on

dynamic co-occurrence networks with a focus on the linkages between keywords. They developed a software named NEViewer to integrate the techniques, algorithms, and measurements in different scientometric methods including co-occurrence, co-authorship and co-citation networks. They employed alluvial diagrams and coloring networks to visualize the trend of changes in the topic at macro and micro scales, respectively [28]. Sitarz et al. [29] proposed a method to identify thematic clusters. As an extended version of another technique developed in 2010 [30], this method constituted two steps. The first step was devoted to article identification, while word processing was achieved in the second step based on which the trend identification could finally be presented. Zhang et al. employed the theories of complex networks to realize a dynamic co-occurrence network in their field of study, i.e. China's urbanization research (CUR). They investigated the correlation of degree, eigenvector and betweenness centrality for networks in different periods. It was found that the betweenness and eigenvector centrality were positively correlated to the degree centrality [31]. Feng et al. combined the co-word analysis with semantic distance to improve the accuracy of analysis results as they believed the two concepts possessed a specific semantic connection [32]. Wang and Chai presented three new indexes for co-word analysis. These indexes were designed to demonstrate the amount of progress in various disciplines, correlation between subjects for networks with different sizes and predict the potential growth points of disciplines [33]. Zhou et al. presented the method of semantic measurement in co-word analysis to tackle the issue of the same amount with different qualities and lack of semantics in correlation calculations. They considered a weight for each obtained attribute, then derived core keywords based on the method of Latent Dirichlet Allocation (LDA). Finally, after forming and transforming into low-dimensional value distributions vectors of top-N keywords, the semantic correlation between keywords was calculated [34]. Katsurai and Ono proposed a technique based on the original dynamic co-word network and obtained sparse dynamic co-word networks. Each network visualized bursty topics over a corresponding period [35]. Finally, Tang and et al. applied the theory of percolation transition with a co-word network for the purpose of filtering out weaker edges [36].

In this article, a co-word analysis is considered in gas turbine maintenance. In this analysis, the words in article titles are used. First, the bibliographic parameters that are determined in HistCite are utilized to define an index. This index is used to determine the critical time periods of the research field, then the networks in each period are built using VOSviewer. By analyzing the networks using SNA in NodeXL, the top hot keywords of each period are found. An overall index is defined which includes four normalized measures, i.e. betweenness centrality, degree centrality, closeness centrality and eigenvector centrality. The value of this index is calculated for each word to determine hotspot keywords. Finally, the results are evaluated using HistCite which is based on author keywords, abstract keywords and title words. These are then aggregated with SNA results to determine the evolutionary trend in the research field.

3 METHODOLOGY, DATA AND ANALYSIS TOOLS

3.1 Methodology

In the last two decades, a number of software have been developed by researchers to facilitate scientometric studies such as VOSviewer, BibExcel, CiteSpace, SCIMAT, UCINET, and HistCite. Some of these are suitable for co-occurrence analysis and evolutionary trend detection, while others are used for visualization in scientometrics. Provided by Eugene Garfield, HistCite is one of the most important software among the mentioned tools. Introduced in 2002, this software was presented as a means of visualizing the history of scientific products [1]. Using this tool, one can analyze the records, authors, words, journals, and organizations, among others, to achieve useful indicators such as total local citation score (TLCS), total global citation score (TGCS), global citation score (GCS) and local citation score (LCS). An example of the implementation of this software in scientometrics can be found in references [37-39]. Another software that has been considered in the field of scientometrics in recent years is VOSviewer, using which bibliometric maps can be displayed graphically [39]. This software has mainly been used to create the maps of journals or authors based on co-citation data or to build keyword maps based on co-occurrence data [40-42]. This computer program uses a technique called visualization of similarities. What is presented in VOSviewer as output is a network of linkages between different terms that, depending on the analysis type, can be a network of connections between different authors, or phrases appearing in titles, keywords or the text of articles and other scientific manuscripts in a research field. So far, this software has been used in scientometrics and knowledge mapping in topics such as building control, public-private partnership, green buildings, and environmental footprint [19, 43-46].

In networks created in VOSviewer, each term forms a node and the connection between terms creates vertices. A node represents an element that can be the name of an author in the co-authorship network or the name of an organization, country or magazine; or maybe subject keywords in the co-word network. To analyze such networks, the concepts employed in social network analysis (SNA), namely the application of graph theory, can be an effective approach despite being rarely addressed until now. The purpose of SNA is to study human relations based on graph theory, while in this research, the objective is to perform this analysis for the relationships between the topics discussed in scientometrics (specifically, the co-word network). So far, SNA has been applied in many disciplines including business and management, computer science, humanities, information science, social sciences, medicine and health [47]. Some of the well-known tools for SNA are NodeXL, VOSviewer, and HistCite. NodeXL allows the user to obtain comprehensive and useful analyses of the network and advantageous indexes from the network without getting involved in the complexities of graph theory. This software was first introduced in 2009 by [48] specifically for SNA, but can generally be used for any other network. In addition, one can calculate different indexes such as network density, network diameter, degree centrality, betweenness centrality, closeness centrality and eigenvector centrality. Among the few uses of NodeXL in scientometrics, one can mention the study by [49] in informatics where the values of betweenness and closeness centrality were calculated for the co-occurrence network.

Generally, each scientometric research is conducted according to a specific methodology. As such, scientometric tools and indexes are used in tandem to achieve the evolutionary trend, knowledge domains, institutions, countries and major journals. In this regard, Si et al. [50] presented a framework in a scientometric analysis that included published journals, countries, regions, institutions, documents co-citation, keyword co-occurrence and cluster analysis for their area of research (bicycle sharing). Furthermore, a comprehensive knowledge map was presented that included knowledge domains, major journals, evolutionary trends, main institutions and countries, and critical papers based on the results of analyses. Using a methodology benefiting from VOSviewer and CiteSpace as analysis tools, Martinez et al. [40] reviewed 1158 articles and provided an outlook of future research in computer vision while presenting a co-word network, a co-authorship network, a network of countries, an author co-citation network and a journal co-citation network.

Fig. 1 shows the methodology used in this research. As observed, the employed procedure is summarized as follows: first, based on the initial analyses performed on the data in HistCite, six major time periods (where the number of citations and the number of records is maximum) in the field of gas turbine maintenance are identified. Next, the data of each period are exported to VOSviewer for analysis where the co-occurrence networks of words are derived. Then, to perform SNA and obtain the centralities, the data are exported to NodeXL via Pajek file format. NodeXL calculates the network centrality scores and overall index based on which the graphs are reproduced. On the other hand, to verify the results, data analysis is performed in HistCite and finally, the common top hot keywords are identified to determine the evolutionary trend.

In this study, the research data are collected from the Web of Science. This database contains valuable contributions from different researchers and is commonly used in scientometric research, either alone or in combination with other databases. The search terms used to extract data from the Web of Science are presented in Tab. 1. Given these search terms, the research into the field of gas turbine design aimed at extending the turbine lifetime or its associated parts, as well as some areas of gas turbine control relevant to the maintenance field, are completely covered.

Table 1 Search terms utilized in the maintenance of gas turbine

ST1	ST2	ST1	ST2
prognos*	gas turbine	condition based monitoring	micro*jet engine
diagnos*	turbo*jet	anomaly detection	small jet engine
fault detection	jet engine	health management	aero*engine
fault detection and identification	turbo*prop	proactive maintenance	marine gas turbine engine
degradation detection	aircraft engine	predictive maintenance	aircraft engine
condition monitoring	gas turbine engine	preventive maintenance	aircraft propulsion

The search terms presented in the table are divided into two categories of ST1 and ST2, and are linked to each other using logical commands in Web of Science. The starred items in the table cover every possible state of a term.

It should be noted that this study only considers articles written in English. The interval for this analysis starts from the time Web of Science first began providing information in this field, i.e. 1985, and ends in December 2019. Fig. 2 shows the number of articles published in the field of gas turbine maintenance from 1985 to 2019. This period is divided into 6 phases: the highest number is held by 2019 with 171 records, whereas the lowest number belongs to the interval 1986-87 with no record. Among these years, six periods are identified and co-word analysis is performed on them. In this research, a combination of the outputs of three software, i.e. HistCite, VOSviewer and NodeXL, is employed to generate co-word networks of different periods and to analyze and evaluate them. For each period, the word network appearing in the title of articles is considered. The reason for this choice is that the phrases in article titles are generally devoid of unnecessary words and refer to the exact same thing the researcher has conducted. Thus, analyzing the networks created based upon these words becomes easier. Furthermore, the network is free of unnecessary words that may appear frequently throughout the article text.

As briefly stated above, the data for each period are first acquired from Web of Science and imported into VOSviewer as input where co-word networks are generated, then the network data is exported to NodeXL for SNA under a file in Pajek format. NodeXL calculates important network indexes such as network density and four centrality scores. In social networks and any network of nodes and edges in general, centrality determines which nodes are effective and of the pivotal role [48]. Different types of centrality include degree, betweenness, closeness and eigenvector centrality. Degree centrality is the first index that demonstrates the importance of a node in the network, and its value, apart from the type and quality of the edge, depends only on the number of edges a node has [48].

Betweenness centrality shows the nodes with a bridge role in the network using a number that resides between 0 and 1 as in other indexes. Thus, if a node has a betweenness score of zero, its removal will not yield anything special in the network. In contrast, if a node with a betweenness centrality of 1 and the role of a bridge between nodes is removed from the network, the communication path between many nodes is cut off [48]. Eigenvector centrality is defined so as to distinguish seemingly (but not truly) valuable nodes in a network from other constructive nodes. A node with a high eigenvector centrality score is connected to nodes that are linked to many other nodes. Sometimes, there are nodes in the network that, although they have a high degree, they themselves have established connections with nodes that have little links with the nodes around them. Such a node is not really of a high value. Thus, eigenvector centrality seeks to detect such nodes in the network [48, 51]. Closeness centrality is the average distance between one node and all other nodes in the network. The nodes located around the network usually have a low closeness centrality score [52, 53]. One should note that some studies have considered this index to be inversely proportional to closeness, while in others, this index is not inverted, meaning that higher centrality indicates less closeness to other nodes in the network. In this article, the second approach is adopted. In other words, close words possess lower closeness centrality scores.

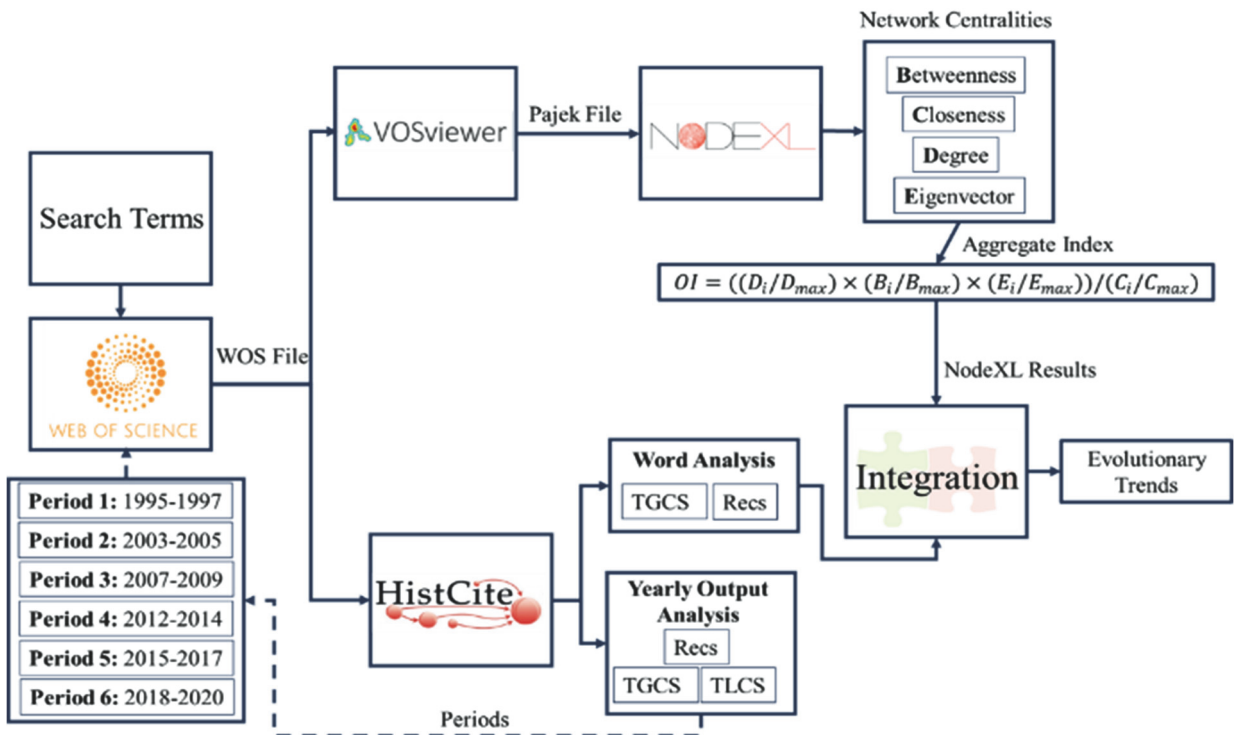


Figure 1 Methodology for determining the evolutionary trend in the field

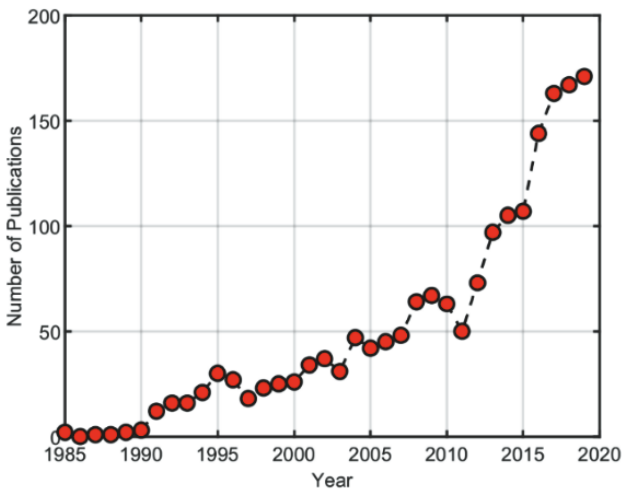


Figure 2 The number of articles published in the field of gas turbine maintenance from 1985 to 2019

There is a special algorithm for calculating each of these indexes that occasionally turns into a complex computational procedure, although NodeXL facilitates this task by performing these calculations in the network. After determining different centrality scores, various networks are regenerated based on the centralities. Here, a measure called overall index (*OI*) is defined and calculated which has the effect of all four network centralities combined. By considering this index, one can examine the trend of researcher's attention to the topic under study in the field of interest to a reasonably acceptable degree. This index contains the four scores mentioned in the network analysis, namely the degree centrality *D*, betweenness centrality *B*, closeness centrality *C*, and eigenvector centrality *E* for each expression *i*, and is defined as:

$$OI = ((D_i/D_{max}) \times (B_i/B_{max}) \times (E_i/E_{max})) / (C_i/C_{max}) \quad (1)$$

where the denominator of fractions represents the maximum value of an index from a set of expressions for a given period. Thus, all indexes are normalized with respect to their maximum values. According to what was expressed in the definition of closeness score, this index has appeared in its inverse form in the above equation. In parallel with the network analyses, Web of Science data is imported into HistCite so that separate analyses are performed based on the indexes defined in this software such as productivity and citation indexes, validating and correcting previous networks if needed. In HistCite analyses, in addition to the terms in titles, the words in article abstracts and author keywords are also used. Finally, all these analyses are presented in the form of a table of words. The analyses in this research are conducted for six different time periods according to the described method. The six mentioned periods are based on the initial analysis performed in HistCite as described in detail in what follows.

3.2 Determining the Periods

In this section, by examining the quality and quantity of articles in HistCite from 1985 to 2020 in the field of gas turbine maintenance, the periods over which the co-word analyses have been performed are obtained. To determine the time periods, the quantity and quality of the articles published in different years are examined. Accordingly, HistCite relies on bibliographic parameters that provide the quantitative and qualitative value of articles. Fig. 3 shows the value of the normalized product of three important bibliographic parameters, namely *TGCS*, *Recs* and *TLCS* in different years which are determined according to:

$$I_{TRT} = \frac{(TLCS \times Recs \times TGCS)}{(TLCS_{max} \times Recs_{max} \times TGCS_{max})} \quad (2)$$

where $TGCS$ represents the total citations in Web of Science to the collection of articles with this exact word in their title, $Recs$ is the number of articles (records) in the collection with this title, and $TLCS$ signifies the total number of citations in the collection to the articles with this word in their title. The parameters with the subscript max in the denominator of the above equation show the maximum value of the index over all these years. Since $TGCS$ and $TLCS$ illustrate the score to citations, these two parameters somehow represent the quality of articles. On the other hand, $Recs$ deals with the number of records and provides the quantitative aspect of articles in each period. These three parameters are determined based on HistCite analyses. In fact, the diagram of Fig. 3 tries to depict all three indicators simultaneously. The chart has an oscillating behavior with increasing amplitude such that the height of bars periodically increases to a maximum and decreases again in some years. In general, however, it has an increasing trend. In this regard, 2016 was at its peak, indicating that the number of highly-cited articles in this year was significant both locally and globally, and that a large number of records were registered that year. In view of this chart and the years 1996, 2004, 2008, 2013 and 2016 as the years that are distinct from their previous and subsequent years, 5 two year periods are defined and the studies conducted in these years are investigated for co-word analysis. For example, the fifth period from 2015 to 2017 has 2016 as one of the five significant years in its middle. The time periods whose co-word analyses are considered in this research includes 1995-1997, 2003-2005, 2007-2009, 2012-2014 and 2015-2017. The middle point of these five periods is distinguished from other intervals by a different color in Fig. 3. In addition to these five periods, the last 2 years (2018 - 2020) are also inspected to cover new studies.

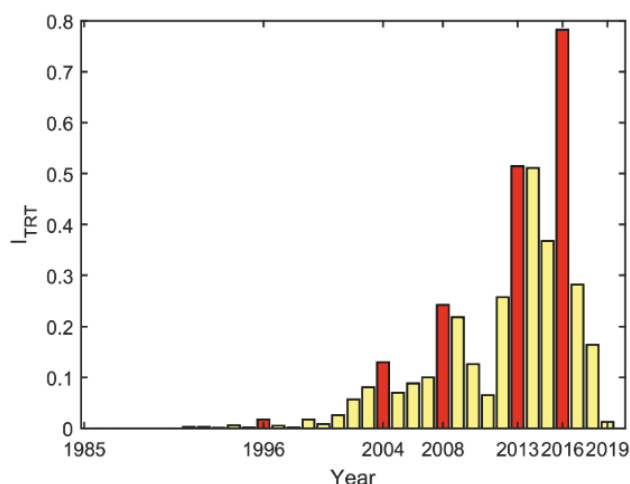


Figure 3 Quality and quantity of articles in terms of I_{TRT} in the field of gas turbine maintenance

4 RESULTS

In this section, the results of co-word analysis are presented for six time periods. Three graphs exist for each period. Each graph displays the value and importance of words over that period in a pair of network centrality measures in a qualitative manner as well as their linkage type. The size of nodes in the graphs is selected according to the degree centrality, betweenness centrality, or

eigenvector centrality. Also, their colors are in accordance with the closeness centrality. To better understand the performed analysis, the hotspot keywords in the first period are described for each of the four indexes along with what can be inferred from each network.

4.1 1st Period (1995-1997)

Fig. 4 shows the degree-closeness co-occurrence network of words appearing in article titles for the period 1995-1997. This network has a density of 0.0689. Darker and larger nodes respectively have greater scores of closeness and degree centrality. This network has a number of nodes with notable degree centrality which respectively include "gas turbine", "application", "design", "development" and "diagnosis". Most of these are general words that are likely to be repeated in the title of most researches in the field of engineering or gas turbine maintenance and are not considered as technical terms in the field under study. However, the connection of these phrases with technical terms can be of considerable importance.

For example, as seen in Fig. 4, the word "design" is connected to the word "gas turbine" which has the highest degree of centrality, and is linked to other terms such as "robust fault detection filter", "unknown input observer", "magnetic bearing", and "integration" which have a high closeness centrality. One also observes that the links between words usually form closed loops. These loops represent a set of the conducted research, and the words on the loop alongside each other generally form article titles. For example, one of the most significant researches in this period in terms of the number of citations is [54] which has linked some of the above words in its title.

In Fig. 5, the size of nodes in the previous network is adjusted proportional to the betweenness centrality, while the color of nodes is based on the closeness index. In addition to phrases like "gas turbine", "diagnosis", "design", "application", other phrases and terms such as "life assessment" and "SPSLIFE" have a high betweenness centrality score. A large value for betweenness centrality in a network indicates that these words have appeared as a potent link between other words, forming connections between different fields. Although the word application in this network has a much larger betweenness score than the last two phrases, these two words are technical terms in the investigated field, thus giving rise to the importance of its links. Since the studied area is the maintenance of gas turbines, it was already expected that terms such as "gas turbine" and "diagnosis" would be prominent in the network. Fig. 6 shows the importance of words in terms of eigenvector centrality in the co-occurrence network. This index displays those nodes that are of higher importance in reality, and utilizing it alongside the degree centrality makes the results of network analysis closer to reality. In this network, the terms "gas turbine" and "application" tangibly possess greater eigenvector centrality score, followed by "diagnosis", "implicit behavioral model", "esprit project TIGER" and "dynamic system" in the center of the graph. This shows these topics have been at the center of attention of researchers.

In this period, the terms "SPSLIFE", "life assessment" and "ceramic components" are among the 10% of phrases

for which the overall index (the indicator presented in the previous section) is maximum. Additionally, word analysis is separately performed in HistCite for this period. According to this analysis, all terms in the title, as well as authors' keywords, are investigated, summing up to a total

of 387 words from 75 records in this period. The importance of the above three words based on the performed analyses in HistCite are shown in Tab. 2. As noticed, all these words are among the top 20% of most-repeated words of this period in the analyses of HistCite

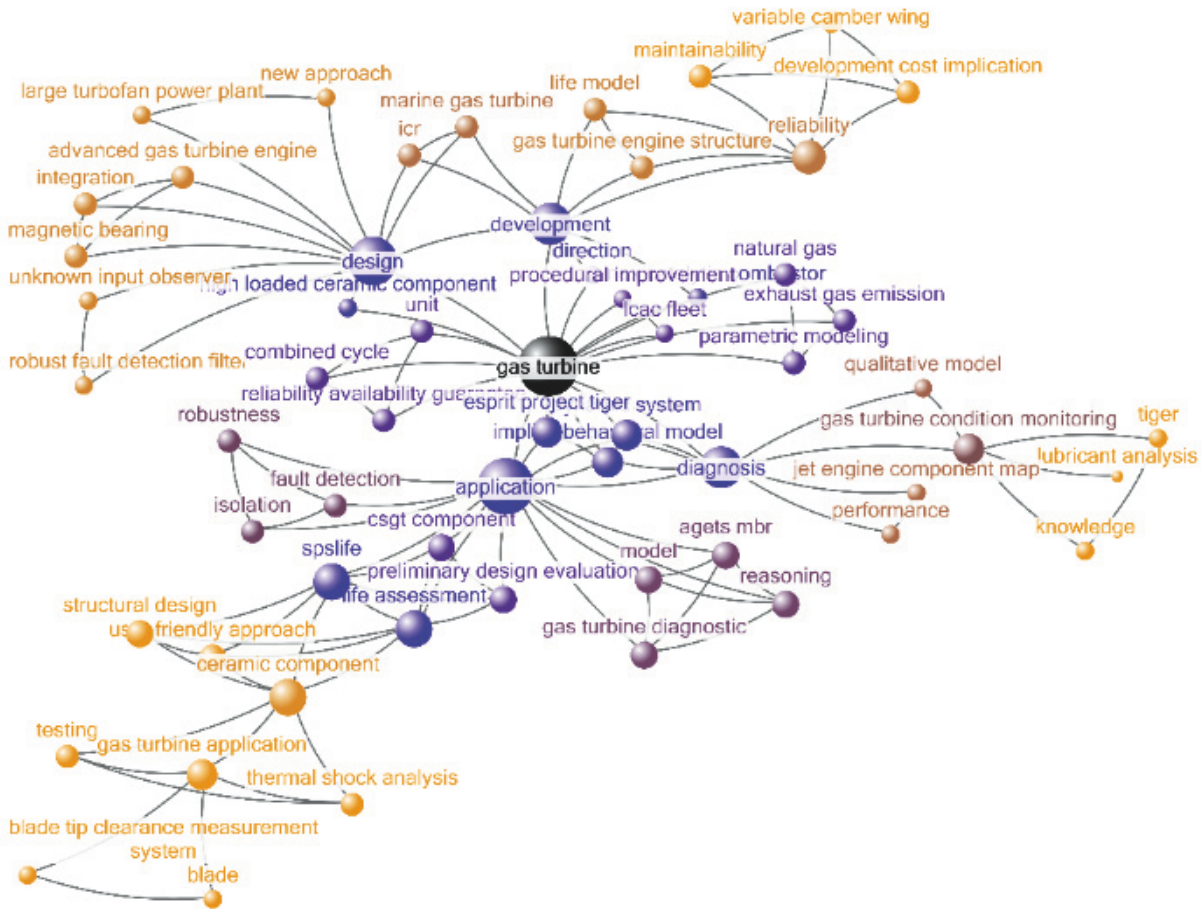


Figure 4 Co-word network in terms of degree-closeness centrality in the field of gas turbine in the first period

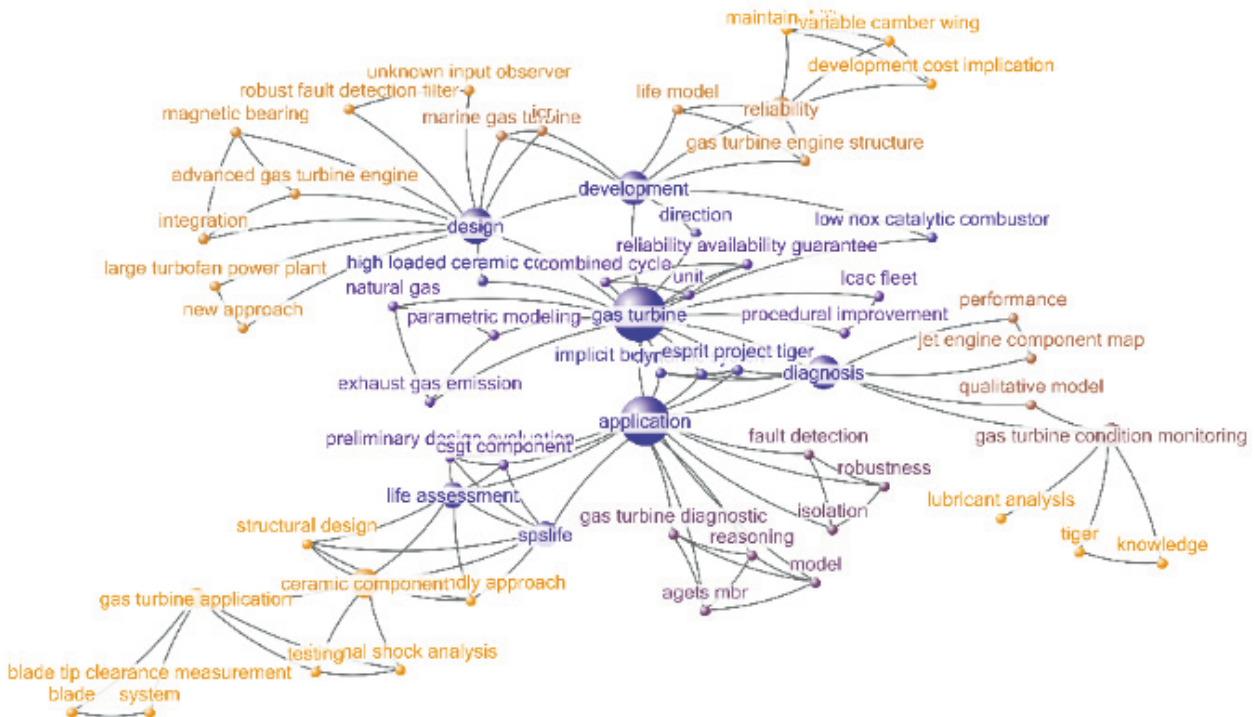


Figure 5 Co-word network in terms of betweenness-closeness centrality in the field of gas turbine maintenance in the first period

combined methods in gas turbine diagnostics that were considered in the fifth period.

According to Tab. 3, the evolutionary trend in the field of gas turbine maintenance for different periods is as follows:

First period (1995-1997): Evaluating the lifetime of ceramic components of gas turbines and assessment of their fatigue life were some of the research topics in this period. Over these years, knowledge-based methods, qualitative fault diagnosis, and observer-based detection and fault isolation systems in the condition monitoring of gas turbines were employed.

Second period (2003-2005): In this period, unstable combustion mechanism, control and prognosis of unstable combustion, online monitoring of combustion and diagnosis of the instability mechanism of periodic combustion by laser were studied. The use of neural networks and Kalman filters in the evaluation of gas turbine performance was considered by researchers, while Kalman filters were used in sensor diagnostics. Furthermore, diagnosis using a limited number of measurements was performed with the aid of combinatorial approaches and optimization techniques. In addition, life cycle assessment and performance analysis of cycles with

reduced carbon dioxide emissions in this period were studied by researchers.

Third period (2007-2009): In this period, Kalman filters were mainly used to detect faults in sensors and actuators while neural networks were employed in performance prognosis, component degradation, estimation of parameters and engine performance in conjunction with filter-based approaches and other performance observers. On the other hand, the performance analysis and techno-economic performance of energy-from-waste combined cycle plants were considered by researchers of this period.

Fourth period (2012-2014): In this period, gas turbine prognosis was achieved using statistical and Monte Carlo method. Reliability analysis was also performed based on different approaches such as Fourier series, response surface and the weakest t-norm, while detecting and isolating the fault were realized using multi-model approaches accompanied by dynamic neural networks. Over these years, thermo-economic, environmental risk and techno-economic analysis along with the discussion of cost optimization for power generation systems that utilize gas turbines in their structure, such as combined cycle power plants and marine propulsion, received considerable attention.

Table 3 Hotspot keywords obtained from SNA of co-occurrence word network in the field of gas turbine maintenance

Period	Word	Degree	Betweenness	Closeness	Eigenvector	Overall index
1995-1997	SPS Life	7	200	59E-4	3174E-5	4084E-5
	Life Assessment	7	200	59E-4	3174E-5	4084E-5
	Ceramic Component	7	270	47E-4	1575E-5	3465E-5
	Gas Turbine Condition Monitoring	5	170	52E-4	986E-5	880E-5
2003-2005	Reliability	6	168	49E-4	642E-5	714E-5
	Combustion Instability	6	2256	26E-4	2178E-5	6181E-5
	Kalman Filter	6	1794	22E-4	1993E-5	5360E-5
	Life Cycle Assessment	8	450	21E-4	2556E-5	2359E-5
	ANN	5	1656	20E-4	412E-5	937E-5
	Measurement	10	1733	23E-4	38E-5	157E-5
2007-2009	Combustion Chamber	17	1433	20E-4	11E-5	72E-5
	Crack	10	2361	20E-4	2483E-5	3438E-5
	ANN	12	1884	19E-4	2104E-5	2959E-5
	Prediction	7	1530	16E-4	858E-5	675E-5
	Kalman Filter	8	966	17E-4	748E-5	408E-5
	Estimation	9	800	15E-4	289E-5	165E-5
	Observer	8	310	15E-4	548E-5	107E-5
2012-2014	Performance Analysis	7	1690	13E-4	38E-5	42E-5
	Waste-to-Energy	7	258	10E-4	8E-5	2E-5
	Biomass	11	7174	5E-4	1243E-5	1754E-5
	Reliability Analysis	12	5568	6E-4	1140E-5	1195E-5
	Statistical Methodology	6	5177	6E-4	699E-5	330E-5
	Reliability	11	6817	5E-4	170E-5	217E-5
	Robust Fault Detection	9	7081	6E-4	171E-5	167E-5
	Techno Economic Analysis	11	1634	5E-4	184E-5	57E-5
2015-2017	Dynamic Neural Network	6	819	5E-4	660E-5	52E-5
	Environment	10	7578	4E-4	30E-5	47E-5
	Clearance	8	3024	4E-4	28E-5	14E-5
	Reliability	39	51709	4E-4	626E-5	4478E-5
	Hybrid	22	18185	4E-4	744E-5	1202E-5
	Prediction	17	17869	4E-4	872E-5	1007E-5
	Swirl	16	15524	4E-4	507E-5	1051E-5
	ANN	17	11680	4E-4	447E-5	357E-5
	Sensor Fault	22	9443	3E-4	237E-5	209E-5
	Prognostic	17	15108	3E-4	166E-5	180E-5
2015-2017	Titanium Alloy	9	4596	3E-4	514E-5	93E-5
	Genetic Algorithm	13	7475	4E-4	209E-5	84E-5
	Isolation	15	3996	3E-4	253E-5	65E-5
	Turbine Blade	7	5497	4E-4	432E-5	65E-5
	Integration	9	6172	3E-4	180E-5	44E-5
	Classification	10	4596	4E-4	233E-5	43E-5
	Biomass	7	3072	3E-4	430E-5	40E-5
	Fusion	17	9245	3E-4	40E-5	31E-5

Table 3 Hotspot keywords obtained from SNA of co-occurrence word network in the field of gas turbine maintenance - continuation

Period	Word	Degree	Betweenness	Closeness	Eigenvector	Overall index
2015-2017	Optimization	15	10364	3E-4	34E-5	26E-5
	Uncertainty	9	12835	3E-4	37E-5	22E-5
	Spray	10	2400	3E-4	194E-5	20E-5
	Erosion	9	4798	3E-4	105E-5	20E-5
	Measurement	10	3592	3E-4	122E-5	20E-5
	Failure Analysis	6	1540	3E-4	495E-5	20E-5
	Life Cycle Assessment	9	3841	3E-4	100E-5	19E-5
	Extreme Learning Machine	7	2533	4E-4	236E-5	17E-5
2018-2020	Useful life Estimation	13	8408	3E-4	25E-5	14E-5
	Reliability	25	67198	4E-4	1719E-5	37708E-5
	Data	29	31109	4E-4	1384E-5	17876E-5
	Economic Analysis	18	6761	3E-4	778E-5	1853E-5
	Multi Objective Optimization	13	14744	3E-4	435E-5	1326E-5
	Useful Life Estimation	11	14042	3E-4	310E-5	762E-5
	Isolation	15	9219	3E-4	139E-5	345E-5
	ANN	14	4371	3E-4	471E-5	336E-5
	Prognostic	13	4869	3E-4	320E-5	503E-5
	Biogas	13	2113	3E-4	998E-5	499E-5
	Performance monitoring	6	19740	3E-4	194E-5	368E-5
	Blade	9	11472	3E-4	180E-5	350E-5
	Availability	8	7535	3E-4	314E-5	294E-5
	Swirl	11	20027	3E-4	71E-5	266E-5
	Useful Life Prediction	14	7191	3E-4	108E-5	206E-5
	Reliability Assessment	12	4609	3E-4	137E-5	144E-5
	Turbine Blisk	7	4362	3E-4	215E-5	111E-5
	Integrated Energy System	7	2703	3E-4	299E-5	93E-5
	Sensor Fault	9	11360	3E-4	30E-5	58E-5
	Signal	8	2190	3E-4	171E-5	54E-5
Big Data	6	1339	3E-4	252E-5	34E-5	

Fifth period (2015-2018): In this period, researchers combined different methods to achieve better systems for the prognosis and diagnosis of faults in the main components of gas turbines and sensors. Some of the methods used for these goals include PCA, PSO-SVM, Fuzzy AHP, Fuzzy TOPSIS, combining gas path analysis with gray relation theory, PF and neural networks. Data fusion-based methods, information fusion and sensor data fusion in performance tracking, degradation modeling and gas turbine fault prognosis were prominently employed over these years. Moreover, instability prediction in combustion chamber with swirl-stabilized combustor, the study of the effects of swirl on flame dynamics and the effects of combustion chamber surface on the flame shape were considered. The topic of prediction in these years was accompanied by addressing different issues such as estimating creep and fatigue life of blades as well as prediction of unsteady turbulent flow, flame, performance, degradation and lifetime. To this end, various tools and methods such as fuzzy methods, neural networks, Bayesian approaches along with other novel techniques were used. Also, detection, isolation and identification of sensor faults based on genetic algorithms, sliding mode observers, networks based on extreme learning machines, deep networks, hybrid Kalman filters and fuzzy models were of special interest to researchers of this period.

Sixth period (2018-2020): In this period, data-driven prognostics with a focus on online monitoring methods, Internet-based monitoring, and the use of dynamic data and big data were considered. Data-level fusion and multi-sensor data fusion methods were exploited in the diagnosis and prognosis, while deep neural networks were used as a means of diagnosis, prognosis and estimation of useful life of components as well as diagnosis of sensors and actuators. Examples of methods used include different types of long-short term memory (LSTM), deep recurrent

neural network (DRNN), bidirectional recurrent neural network (BRNN), convolutional neural network (CNN) and deep belief network (DBN). In this period, atomization, vortex, viscosity effect and some other fluid studies related to swirl injectors in different conditions, especially for new injectors in combustion chamber topics, were investigated by researchers. Also, fatigue life analysis, analysis of failure modes and surface roughness, and reliability assessment for blisk were performed.

Accordingly, the hotspot keywords in 6 periods were categorized as follow:

- 1995-1997: SPS Life, Life Assessment, Ceramic Component.
- 2003-2005: Combustion Instability, Kalman Filter, Life Cycle Assessment, ANN, Measurement, Combustion Chamber.
- 2007 to 2009: Crack, ANN, Prediction, Estimation, Observer, Waste-to-Energy, Kalman Filter.
- 2012 - 2014: Biomass, Reliability Analysis, Statistical Methodology, Robust Fault Detection, Techno Economic Analysis, Dynamic Neural Network, Environment, Clearance.
- 2015-2017: Hybrid, Prediction, Swirl, ANN, Sensor Fault, Prognostic, Genetic Algorithm, Isolation, Turbine Blade, Classification, Integration, Biomass, Fusion, Optimization, Uncertainty, Erosion, Measurement, Life Cycle Assessment, Extreme Learning Machine.
- 2018-2020: Data, Economic Analysis, Multi Objective Optimization, Useful Life Estimation, Isolation, ANN, Prognostic, Biogas, Blade Availability, Swirl, Turbine Blisk, Sensor Fault, Signal, Big Data.

The output of such an analysis was deriving phrases and keywords that caught the attention of researchers more than others in each period. These expressions were then used as the basis of presenting the evolutionary trend in the

field of gas turbine maintenance in the studied periods. For example, the scientometric results of this study allow us to identify areas including combustion chamber diagnosis, sensor monitoring, economic analysis, estimation of turbine's remaining useful life, blade monitoring and turbine blisk reliability estimation from the prominent research of recent years as presented in Fig. 7.

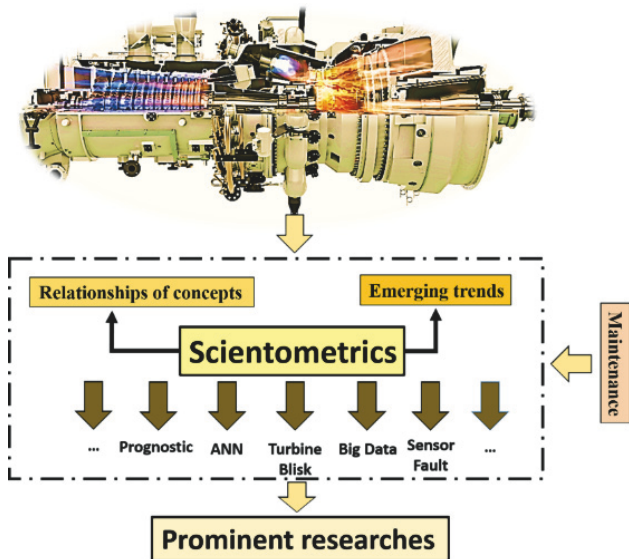


Figure 7 Hotspot topics and prominent researches in gas turbine maintenance resulting from scientometric analysis

5 CONCLUSIONS

The current article presented an investigation into the scientometrics of gas turbine maintenance via co-word analysis in combination with social network analysis. To this aim, the required data were collected from the Web of Science, and well-known tools of scientometrics, generation and analysis of networks were employed. The analyses were performed for six different periods chosen from 1985 to 2020. When selecting these years, *TGCS*, *Recs* and *TLCS* parameters were considered for each year. The years in which these parameters reached their maximums became the basis for determining the periods. Next, four network centrality measures were calculated for the co-word networks of these periods and for each word, an overall index containing all four centrality measures was presented. Technical words and phrases that had the maximum overall network index were presented as hot keywords of each period. They were then validated using the results of HistCite analyses.

Categorized phrases in different periods can be exploited as a reference for researchers in the field of gas turbine maintenance. When the time period of a phrase is known, one can identify the most notable studies of researchers in that period by referring to the records. In addition, one can search for the emergence time of a specific keyword in its earlier years. Accordingly, one can understand the time of rise and decline of different methods and topics or acquire a vision of the future in the studied field. As a suggestion, the analytical method presented in this research can be used in other scientific areas.

6 REFERENCES

- [1] Garfield, E. (2009). From the science of science to Scientometrics visualizing the history of science with HistCite software. *Journal of Informetrics*, 3(3), 173-179. <https://doi.org/10.1016/j.joi.2009.03.009>
- [2] Tijssen, R. (1993). A scientometric cognitive study of neural network research: expert mental maps versus bibliometric maps. *Scientometrics*, 28(1), 111-136. <https://doi.org/10.1007/bf02016288>
- [3] Montoya, F. G., Montoya, M. G., Gomez, J., Manzano-Agugliaro, F., & Alameda-Hernandez, E. (2014). The research on energy in Spain: A scientometric approach. *Renewable and Sustainable Energy Reviews*, 29, 173-183. <https://doi.org/10.1016/j.rser.2013.08.094>
- [4] Kumar, S. & Garg, K. C. (2005). Scientometrics of computer science research in India and China. *Scientometrics*, 64(2), 121-132. <https://doi.org/10.1007/s11192-005-0244-9>
- [5] He, T., Zhang, J., & Teng, L. (2005). Basic research in biochemistry and molecular biology in China: A bibliometric analysis. *Scientometrics*, 62(2), 249-259. <https://doi.org/10.1007/s11192-005-0018-4>
- [6] Akhavan, P., Ebrahim, N. A., Fetрати, M. A., & Pezeshkan, A. (2016). Major trends in knowledge management research: a bibliometric study. *Scientometrics*, 107(3), 1249-1264. <https://doi.org/10.1007/s11192-016-1938-x>
- [7] Mokhtari, H., Barkhan, S., Haseli, D., & Saberi, M. K. (2020). A bibliometric analysis and visualization of the Journal of Documentation: 1945-2018. *Journal of Documentation*. <https://doi.org/10.1108/JD-08-2019-0165>
- [8] Norris, M. & Oppenheim, C. (2010). The h-index: A broad review of a new bibliometric indicator. *Journal of Documentation*. <https://doi.org/10.1108/00220411011066790>
- [9] Lu, F., Ju, H., & Huang, J. (2016). An improved extended Kalman filter with inequality constraints for gas turbine engine health monitoring. *Aerospace Science and Technology*, 58, 36-47. <https://doi.org/10.1016/j.ast.2016.08.008>
- [10] Mohammadi, E. & Montazeri-Gh, M. (2015). A fuzzy-based gas turbine fault detection and identification system for full and part-load performance deterioration. *Aerospace Science and Technology*, 46, 82-93. <https://doi.org/10.1016/j.ast.2015.07.002>
- [11] Li, R., Verhagen, W. J., & Curran, R. (2020). Toward a methodology of requirements definition for prognostics and health management system to support aircraft predictive maintenance. *Aerospace Science and Technology*, 105877. <https://doi.org/10.1016/j.ast.2020.105877>
- [12] Fei, C. W., Tang, W. Z., & Bai, G. C. (2014). Novel method and model for dynamic reliability optimal design of turbine blade deformation. *Aerospace Science and Technology*, 39, 588-595. <https://doi.org/10.1016/j.ast.2014.07.003>
- [13] Mohammadi, E. & Montazeri-Gh, M. (2016). Active fault tolerant control with self-enrichment capability for gas turbine engines. *Aerospace Science and Technology*, 56, 70-89. <https://doi.org/10.1016/j.ast.2016.07.003>
- [14] Chen, X., Zheng, H., Pan, G., & Jia, X. (2014). Parametric modelling system of gas turbine combustor. *Tehnički vjesnik*, 21(6), 1213-1219.
- [15] Direk, M. & Mert, M. S. (2018). Comparative and exergetic study of a gas turbine system with inlet air cooling. *Tehnički vjesnik*, 25(2), 306-311. <https://doi.org/10.17559/TV-20160811162110>
- [16] Montazeri-Gh, M., Fashandi, S. A. M., & Jafari, S. (2018). Theoretical and experimental study of a micro jet engine start-up behaviour. *Tehnički vjesnik*, 25(3), 839-845. <https://doi.org/10.17559/TV-20170316192008>
- [17] Cambrosio, A., Limoges, C., Courtial, J. P., & Laville, F. (1993). Historical scientometrics? Mapping over 70 years of biological safety research with cword analysis. *Scientometrics*, 27(2), 119-143.

- <https://doi.org/10.1007/BF02016546>
- [18] Ravikumar, S., Agrahari, A., & Singh, S. (2015). Mapping the intellectual structure of scientometrics: A co-word analysis of the journal *Scientometrics* (2005-2010). *Scientometrics*, 102(1), 929-955. <https://doi.org/10.1007/s11192-014-1402-8>
- [19] Wuni, I. Y., Shen, G. Q., & Osei-Kyei, R. (2019). Scientometric review of global research trends on green buildings in construction journals from 1992 to 2018. *Energy and Buildings*. <https://doi.org/10.1016/j.enbuild.2019.02.010>
- [20] Shen, L., Xiong, B., & Hu, J. (2017). Research status, hotspots and trends for information behavior in China using bibliometric and co-word analysis. *Journal of Documentation*. <https://doi.org/10.1108/JD-10-2016-0125>
- [21] Yue, H. (2012). Mapping the intellectual structure by co-word: A case of international management science. *International Conference on Web Information Systems and Mining*, 621-628. https://doi.org/10.1007/978-3-642-33469-6_77
- [22] Teichert, T., Heyer, G., Schöntag, K., & Mairif, P. (2011). Co-word analysis for assessing consumer associations: a case study in market research. *Affective Computing and Sentiment Analysis*, 115-124. https://doi.org/10.1007/978-94-007-1757-2_10
- [23] Law, J., Bauin, S., Courtial, J., & Whittaker, J. (1988). Policy and the mapping of scientific change: A co-word analysis of research into environmental acidification. *Scientometrics*, 14(3-4), 251-264. <https://doi.org/10.1007/BF02020078>
- [24] Wu, Y., Jin, X., & Xue, Y. (2017). Evaluation of research topic evolution in psychiatry using co-word analysis. *Medicine*, 96(25). <https://doi.org/10.1097/MD.00000000000007349>
- [25] Hou, Q., Mao, G., Zhao, L., Du, H., & Zuo, J. (2015). Mapping the scientific research on life cycle assessment: a bibliometric analysis. *The International Journal of Life Cycle Assessment*, 20(4), 541-555. <https://doi.org/10.1007/s11367-015-0846-2>
- [26] Xiao, Q. (2016). Node importance measure for scientific research collaboration from hypernetwork perspective. *Tehnički vjesnik*, 23(2), 397-404. <https://doi.org/10.17559/TV-20150314110749>
- [27] Zhang, W., Zhang, Q., Yu, B., & Zhao, L. (2015). Knowledge map of creativity research based on keywords network and co-word analysis, 1992-2011. *Quality & Quantity*, 49(3), 1023-1038. <https://doi.org/10.1007/s11135-014-0032-9>
- [28] Wang, X., Cheng, Q., & Lu, W. (2014). Analyzing evolution of research topics with NEViewer: a new method based on dynamic co-word networks. *Scientometrics*, 101(2), 1253-1271. <https://doi.org/10.1007/s11192-014-1347-y>
- [29] Sitarz, R., Heneczowski, M., Jabłońska-Sabuka, M., & Krasławski, A. (2015). Clustering Method for Analysis of Research Fields: Examples of Composites, Nanocomposites and Blends. *Intelligent Systems' 2014*, 431-442. https://doi.org/10.1007/978-3-319-11310-4_37
- [30] Sitarz, R., Krasławski, A., & Jezowski, J. (2010). Dynamics of knowledge flow in research on distillation. *Computer Aided Chemical Engineering*, 28, 583-588. [https://doi.org/10.1016/S1570-7946\(10\)28098-7](https://doi.org/10.1016/S1570-7946(10)28098-7)
- [31] Zhang, Q. R., Li, Y., Liu, J. S., Chen, Y. D., & Chai, L. H. (2017). A dynamic co-word network-related approach on the evolution of China's urbanization research. *Scientometrics*, 111(3), 1623-1642. <https://doi.org/10.1007/s11192-017-2314-1>
- [32] Feng, J., Zhang, Y. Q., & Zhang, H. (2017). Improving the co-word analysis method based on semantic distance. *Scientometrics*, 111(3), 1521-1531. <https://doi.org/10.1007/s11192-017-2286-1>
- [33] Wang, M. & Chai, L. (2018). Three new bibliometric indicators/approaches derived from keyword analysis. *Scientometrics*, 116(2), 721-750. <https://doi.org/10.1007/s11192-018-2768-9>
- [34] Zhou, L., Ba, Z., Fan, H., & Zhang, B. (2018). Research on the semantic measurement in co-word analysis. *International conference on information*, 409-419. https://doi.org/10.1007/978-3-319-78105-1_45
- [35] Katsurai, M. & Ono, S. (2019). TrendNets: mapping emerging research trends from dynamic co-word networks via sparse representation. *Scientometrics*, 121(3), 1583-1598. <https://doi.org/10.1007/s11192-019-03241-6>
- [36] Tang, M. C., Teng, W., & Lin, M. (2019). Determining the critical thresholds for co-word network based on the theory of percolation transition. *Journal of Documentation*. <https://doi.org/10.1108/JD-06-2019-0117>
- [37] Bornmann, L. & Marx, W. (2012). HistCite analysis of papers constituting the h index research front. *Journal of Informetrics*, 6(2), 285-288. <https://doi.org/10.1016/j.joi.2011.11.001>
- [38] Mohammadi, S. J., Miran Fashandi, S. A., Jafari, S., & Nikolaidis, T. (2020). A scientometric analysis and critical review of gas turbine aero-engines control: From Whittle engine to more-electric propulsion. *Measurement and Control*. <https://doi.org/10.1177/0020294020956675>
- [39] Tho, S. W., Yeung, Y. Y., Wei, R., Chan, K. W., & So, W. W. M. (2017). A systematic review of remote laboratory work in science education with the support of visualizing its structure through the histcite and citespace software. *International Journal of Science and Mathematics Education*, 15(7), 1217-1236. <https://doi.org/10.1007/s10763-016-9740-z>
- [40] Martinez, P., Al-Hussein, M., & Ahmad, R. (2019). A scientometric analysis and critical review of computer vision applications for construction. *Automation in Construction*, 107, 102947, 2019. <https://doi.org/10.1016/j.autcon.2019.102947>
- [41] Van Eck, N. J. & Waltman, L. (2017). Citation-based clustering of publications using CitNetExplorer and VOSviewer. *Scientometrics*, 111(2), 1053-1070. <https://doi.org/10.1007/s11192-017-2300-7>
- [42] Xie, L., Chen, Z., Wang, H., Zheng, C., & Jiang, J. (2020). Bibliometric and Visualized Analysis of Scientific Publications on Atlantoaxial Spine Surgery Based on Web of Science and VOSviewer. *World Neurosurgery*. <https://doi.org/10.1016/j.wneu.2020.01.171>
- [43] Martinez, S., del Mar Delgado, M., Marin, R. M., & Alvarez, S. (2019). Science mapping on the Environmental Footprint: A scientometric analysis-based review. *Ecological Indicators*, 106, 105543. <https://doi.org/10.1016/j.ecolind.2019.105543>
- [44] Park, J. Y. & Nagy, Z. (2018). Comprehensive analysis of the relationship between thermal comfort and building control research-A data-driven literature review. *Renewable and Sustainable Energy Reviews*, 82, 2664-2679. <https://doi.org/10.1016/j.rser.2017.09.102>
- [45] Song, J., Zhang, H., & Dong, W. (2016). A review of emerging trends in global PPP research: analysis and visualization. *Scientometrics*, 107(3), 1111-1147. <https://doi.org/10.1007/s11192-016-1918-1>
- [46] Wang, Q. (2018). Distribution features and intellectual structures of digital humanities. *Journal of Documentation*. <https://doi.org/10.1108/JD-05-2017-0076>
- [47] Schultz-Jones, B. (2009). Examining information behavior through social networks. *Journal of Documentation*. <https://doi.org/10.1108/00220410910970276>
- [48] Hansen, D., Shneiderman, B., & Smith, M. A. (2010). Analyzing social media networks with NodeXL: Insights from a connected world. Morgan Kaufmann. <https://doi.org/10.1016/B978-0-12-382229-1.00002-3>
- [49] Sedighi, M. (2016). Application of word co-occurrence analysis method in mapping of the scientific fields (case study: the field of Informetrics). *Library Review*. <https://doi.org/10.1108/LR-07-2015-0075>

- [50] Si, H., Shi, J. G., Wu, G., Chen, J., & Zhao, X. (2019). Mapping the bike sharing research published from 2010 to 2018: A scientometric review. *Journal of cleaner production*, 213, 415-427. <https://doi.org/10.1016/j.jclepro.2018.12.157>
- [51] Tsvetovat, M. & Kouznetsov, A. (2011). *Social Network Analysis for Startups: Finding connections on the social web*. O'Reilly Media, Inc.
- [52] Wasserman, S. & Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge university press. <https://doi.org/10.1017/CBO9780511815478>
- [53] Maiya, A. S. & Berger-Wolf, T. Y. (2010). Online sampling of high centrality individuals in social networks. *Pacific-Asia Conference on Knowledge Discovery and Data Mining*, 91-98. https://doi.org/10.1007/978-3-642-13657-3_12
- [54] Chen, J., Patton, R. J., & Zhang, H. Y. (1996). Design of unknown input observers and robust fault detection filters. *International Journal of control*, 63(1), 85-105. <https://doi.org/10.1080/00207179608921833>

Contact information:**Ali NEKOONAM**, PhD

Department of Mechanical Engineering,
Iran University of Science and Technology,
Tehran, Iran
E-mail: ali_nekoonam@mecheng.iust.ac.ir

Reza Fatehi NASAB

Department of Mechanical Engineering,
Iran University of Science and Technology,
Tehran, Iran
Email: r.fatehinasab@gmail.com

Soheil JAFARI, PhD

Centre of Propulsion Engineering, Cranfield University,
Cranfield, UK
E-mail: S.Jafari@cranfield.ac.uk

Theoklis NIKOLAIDIS, PhD

Centre of Propulsion Engineering, Cranfield University,
Cranfield, UK
E-mail: t.nikolaidis@cranfield.ac.uk

Nader ALE EBRAHIM, Profesor

Office of the Deputy Vice-Chancellor (Research & Innovation),
University of Malaya,
Jalan Lembah Pantai,
Kuala Lumpur, Wilayah Persekutuan, 50603, Malaysia
E-mail: aleebrahim@perdana.um.edu.my

Seyed Alireza MIRAN FASHANDI, PhD

(Corresponding author)
Department of Mechanical Engineering,
Iran University of Science and Technology,
Tehran, Iran
E-mail: s.alireza.miran@gmail.com

Abbreviation	Description
SNA	Social network analysis
TLCS	total local citation score
TGCS	total global citation score
GCS	global citation score
LCS	local citation score
SPS	Special protection systems
ANN	Artificial neural network
DNN	dynamic neural network
Blisk	Blade integrateddisk
PCA	Principal component analysis
PF	Particle filter

2023-01-31

A scientometric methodology based on co-word analysis in gas turbine maintenance

Nekoonam, Ali

Faculty of Mechanical Engineering in Slavonski Brod. et al.

Nekoonam A, Naseb RF, Jafari S, et al., (2023) A scientometric methodology based on co-word analysis in gas turbine maintenance. Tehnicki Vjesnik/Technical Gazette, Volume 30, Issue 1, January 2023, 361-372

<https://doi.org/10.17559/TV-20220118165828>

Downloaded from Cranfield Library Services E-Repository