https://doi.org/10.1016/j.jmsy.2021.11.013

Thematic analysis of sustainable ultra-precision machining by using text mining and unsupervised learning method

Hongting Zhou[1], Wai Sze Yip[1], Jingzheng Ren[2] and Suet To[1*]

[1]State Key Laboratory in Ultra-precision Machining Technology, Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China

[2]Department of Industrial and Systems Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong SAR, China

*Corresponding author: Suet To. E-mail address: sandy.to@polyu.edu.hk

Abstract

Sustainable manufacturing is one key research area to reduce environmental damages and resource waste nowadays. As a cutting-edge manufacturing method, ultra-precision machining (UPM) plays an increasingly significant role to achieve sustainable manufacturing because of its rapidly increasing demand. The purpose of this paper is to discover and evaluate the main themes of current works about sustainable UPM. By utilizing the latent Dirichlet allocation (LDA) method to analyze the abstracts of the relevant publications, four main themes of sustainable UPM were identified. The percentage of each documents' content contributing to these four themes was also extracted. According to the documents' contribution data, the publications can be classified into four groups by using the K-means algorithm. It shows that the machining process is the most focused theme in this field and the majority of works about surface structure involved multiple topics. And the social aspect of sustainable UPM needs extensive investigation in the future. In this paper, the thematic analysis was conducted for the first time in the area of sustainable UPM. And the LDA-unpreserved learning approach was also proposed in this work originally. This work provides an overall map of sustainable UPM literature to help researchers select the topics which have not been discussed.

Keywords: Sustainable development; Ultra-precision machining; Thematic analysis; Text mining; Latent Dirichlet allocation; Unsupervised learning

1. Introduction

Sustainable development has been considered as one of the most important aspects in the manufacturing industry to reduce environmental impact and improve the efficiency of resource utilization (Schneider, Das, Kirsch, Linke, & Aurich, 2019). One main objective of Industry 4.0 is reducing resource consumption and environmental damage (Darwish, El-Wakad, & Farag, 2021). Various parameters like material, technology, energy, social impact need to be considered by researchers to achieve sustainable manufacturing (Ocampo-Martinez & Olaru, 2020; Stoycheva et al., 2018). As pointed out by Schneider et al. (2019), the lack of study in some aspects, such as organizational aspects and embodied energy, may also bring barriers to achieving sustainable UPM. Therefore, sustainable manufacturing is still one of the main challenges in both industry and academia. Ultra-precision machining (UPM) is a cutting-edge manufacturing technology, which can achieve high surface accuracy and complicated geometry (W. Yip, To, & Zhou, 2020). As the economic growth, the demand for UPM products has experienced a rapid increase in the manufacturing sector (Gao & Huang, 2017). For instance, UPM is necessary to produce high-precision components of computers and medical devices (Corbett, McKeown, Peggs, & Whatmore, 2000). The global market size of the ultra-precision machine was predicted to reach USD 528.9 million by 2026, from USD 432.8 million in 2020 (More, 2020). Therefore, UPM has also become a remarkably significant topic to achieve sustainable development.

Currently, a few works have been done to improve sustainability in UPM, which involves different concepts, issues, and methods like cutting experiments (W. Yip & To, 2020), and sustainability assessment (Gupta et al., 2020). The main challenges and opportunities of sustainable UPM have also been reviewed by W. S. Yip, To, and Zhou (2021). To summarize the existed research work and predict the developing trend, a review paper of sustainable UPM has been published by Schneider et al. (2019). However, there are still several research gaps in this area:

1. It still needs a model to identify the main concepts and themes in sustainable UPM literature, which can help to summarize and present the subfields of current research (De la Hoz-Correa, Muñoz-Leiva, & Bakucz, 2018).
2. It still lacks a method to classify the current works of sustainable UPM based on the themes they belong to, which can show their distribution among the thematic map (Jalali, Razak, Gordon, Perakslis, & Madnick, 2019).
3. And the overall implementations based on subfields of sustainable UPM still need to be delivered based on the status of the main subfields.

To address the above limitations, a thematic analysis model based on text mining and unsupervised learning methods is proposed in this paper. Firstly, the abstracts of the sustainable UPM literature were searched and downloaded from the Web of Science. Then, the bibliometric analysis of these documents was conducted by using VOSviewer. As latent Dirichlet allocation (LDA) can extract the latent topics from a collection of

textual documents (Bastani, Namavari, & Shaffer, 2019), it was used to discover the main themes among these publications. Besides, the percentage of content that belongs to each theme can also be detected by this model, which can show the contributions of one publication to every theme. Moreover, the documents can be classified into several clusters based on documents' contribution data by using the K-means algorithm, which is an unsupervised learning method suitable for the classification of a small dataset (Luo et al., 2018). To plot the clustering map, principal component analysis (PCA) was utilized to reduce the dimension of contribution data. In this way, the main streams of sustainable UPM can be identified and evaluated to give more implementations in this field.

2. Theory

2.1. Thematic analysis

Thematic analysis is defined as a method to identify, analyze, and discover the patterns about themes for a set of documents (Braun & Clarke, 2006). After selecting the data relevant to the objective of the project, it requires encoding the data to uncover the meaningful themes and concepts of one area (Ali, Mustapha, Osman, & Hassan, 2020). This method has been applied in various research areas to figure out the characteristic of the focused topics in specific fields. For example, the research works about the social sustainability of supply chain firms have been evaluated by using thematic analysis to discover the main research branches (Sodhi & Tang, 2018). And a thematic analysis of international entrepreneurship research has also been conducted to generate the thematic map, and domain ontology (Jones, Coviello, & Tang, 2011). In the field of sustainable manufacturing, thematic analysis has been utilized to evaluate the relationship between the concept of lean and sustainable manufacturing (Hartini & Ciptomulyono, 2015). However, the thematic analysis of the research work about sustainable UPM has not been finished.

2.2. Bibliometric analysis

Bibliometric analysis is a method to evaluate bibliographic material by quantitative methods, which is originated from information science (BROADUS, 1987). It has attracted the attention of numerous researchers in multiple research fields. For example, the research of COVID-19 has been analyzed by comparing the publication in different regions (Fan et al., 2020). Moreover, Maditati, Munim, Schramm, and Kummer (2018) have utilized bibliometric analysis to discover the trend of green supply chain management. And the main research categories of renewable energy finance have been summarized by bibliometric analysis (Elie, Granier, & Rigot, 2020). In the research area of manufacturing, this approach has been applied to investigate lean manufacturing (De Oliveira, Sousa, & De Campos, 2019), and additive manufacturing (Caviggioli & Ughetto, 2019).

2.3. Latent Dirichlet allocation

LDA is a hierarchical Bayesian model to conduct topic modeling, which focuses on discovering main topics (themes) from multinomial documents (Xiong, Cheng, Zhao, & Liu, 2019). After being introduced by Blei, Ng, and Jordan (2003), it shows its potential in various research areas involving textual data. For instance, an LDA-based model has been developed for financial time series prediction by Kanungsukkasem and Leelanupab (2019). Besides, it has been applied in the transportation area to discover the trend and main topics (Sun & Yin, 2017). What's more, LDA has played a significant role in investigating the development of multi-tier supply chains (R. Zhou, Awasthi, & Stal-Le Cardinal, 2020), and the trends of heavy-duty electric battery vehicles research (Naumanen, Uusitalo, Huttunen-Saarivirta, & van der Have, 2019). But it still lacks the applications of LDA in sustainable UPM literature.

2.4. Unsupervised learning

Machine learning has shown its effectiveness to do forecasting and classification in multiple areas like image processing (Z. Wang, Zhang, & Zhou, 2019), demand forecasting (H. Zhou, Yip, Ren, & To, 2020), failure prediction (Leukel, González, & Riekert, 2021), and reconfigurable manufacturing (Morgan, Halton, Qiao, & Breslin, 2021). Unsupervised learning refers to the type of machine learning algorithms dealing with unlabeled data to do classification and dimension reduction (Kim et al., 2019), which includes the k-means algorithm and PCA. And K-means has been one of the most used classification algorithms with applications in the emission control system (Pacella, 2018), energy clusters determination (Marino & Marufuzzaman, 2020). Besides, a cluster analysis was conducted for the segment households in a Portuguese urban area by using the K-means algorithm (Vicente & Reis, 2007). Moreover, China's industry sectors have also been assessed based on the K-means algorithm by Gengyuan Liu, Yang, Hao, and Zhang (2018). However, the K-means method has not been applied to classify the work in the UPM field.

For PCA, it was pointed as a multivariate technique to transform the original data set to a new one by linear combinations (X. Chen, Samson, Tocqueville, & Aubin, 2015), which is widely used in dimension reduction (Shi & Tsung, 2003). For instance, a production scheduling decision support system has been developed based on PCA and logistic regression (Mehrjoo & Bashiri, 2013). And PCA has also been utilized to predict the laser welding status by Guiqian Liu, Gao, You, and Zhang (2019). Also, a tool wear monitoring model has been established according to PCA (G. Wang, Zhang, Liu, Xie, & Xu, 2019). What's more, a sustainability assessment method based on PCA was applied to evaluate the sustainable performance of firms (Jiang et al., 2018). And Han, Mannan, Stein, Pattipati, and Bollas (2021) have developed a machine state prediction system

based on the PCA method. But it still lacks the work to apply PCA into the evaluation of the sustainable UPM literature.

3. Methodology

3.1. Data collection

Web of Science was utilized to find the works about the sustainable UPM. The papers were searched by the term "sustainable ultra-precision machining" and its synonyms like "sustainable micro-manufacturing". As a result, a total of 143 relevant publications was found, including journal papers, conference papers, and books. And these documents were also filtered manually to exclude the publications which are not relevant to UPM as the data size is small. After that, only 106 publications were selected to form the raw dataset, which is shown in **Supplemental Material**. Then, the abstracts of these publications were downloaded to conduct the thematic analysis. As the data size is quite small, these papers were not divided into different periods according to their publication years.

3.2. Bibliometric analysis

In this paper, the VOSviewer was used to conduct the bibliometric analysis to provide an overall concept of the research area of sustainable UPM. VOSviewer is a computer program to conduct bibliometric analysis, which can provide a graphical representation of bibliometric maps (Van Eck & Waltman, 2010). The selected publications in the data collection process were input to this software to generate the author citation network, source citation network, and country citation network. Besides, the authors and countries/regions have been ranked according to the citation numbers by VOSviewer to see which author and region have the most impact on the research field of sustainable UPM. Moreover, the classification of keywords based on co-occurrence of keywords has also been conducted to compare with the LDA method used in this work to do the model validation in section 4.4.

3.3. Principle of the latent Dirichlet allocation

The graphical representation of LDA is shown in **Figure 1**. As described by Xiong et al. (2019), the total number of documents is donated M and the total number of words in a document is N, while K is the number of topics (themes). Firstly, the topic-word distribution φ_k is sampled from the prior distribution Dirichlet(β). For each document $m \in M$, the document length N_m is sampled from $Poiss(\xi)$ and the document-topic distribution θ_m from the prior distribution Dirichlet(α). And for each word $w_{m,n}$ in document m, the topic index $z_{m,n}$ is sampled from the document-topic distribution $Mult(\theta_m)$, and the word index $w_{m,n}$ is sampled

from the topic-word distribution $Mult(\varphi_{z_{m,n}})$. Besides, to determine the optimal number of topics K, the coherence score has been proposed to explore the coherence of a topic (Syed & Spruit, 2017). The core idea of this measurement is that words with similar meanings tend to occur in similar contexts (Aletras & Stevenson, 2013). In this work, the coherence measurement was utilized to find the optimal number of themes of the sustainable UPM literature. By setting the themes number enabling the coherence score to reach the peak, it can prevent the underfitting or overfitting for the LDA model (Helal & Mouhoub, 2018). In this paper, the genism package of Python was used to build up the LDA model.

In this method, the documents are assumed as random mixtures over latent topics (Q. Chen, Yao, & Yang, 2016). And the texts are represented as a "bag of words", in which the order of words, syntax, and punctuation are ignored. And the main final output of the LDA model is topic-word distributions (Xiong et al., 2019), in which each word has probabilities for belonging to each topic (or theme). Only the top ten keywords of each theme with the highest probabilities were selected in this paper. And topic combinations are iterated during the training process of the model (Shiryaev, Dorofeev, Fedorov, Gagarina, & Zaycev, 2017). Therefore, the main function of this model is allocating the keywords to present the latent topics (or themes) instead of creating themes.

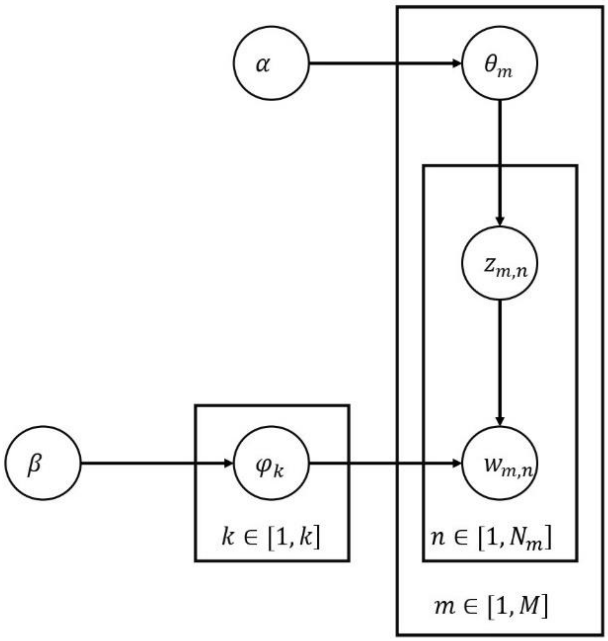

Figure 1 The graphical representation of LDA

3.4. K-means for documents classification

The publications could be classified by using the K-means algorithm efficiently based on the document-topics distribution data. This algorithm can classify the data into k clusters based on the Euclidean Distance (Li, Chen, Chen, Chen, & Shen, 2020):

$$d_{ij} = (\sum_{l=1}^{p}(x_{il} - c_{kl})^2)^{1/2} \quad (1)$$

where d_{ij} is the distance between document i and the center of the cluster k; and x_{il} stands for the contribution percentage to theme l from document i; and c_{kl} is the value of theme l for the center of cluster k. As described by Hartigan and Wong (1979), k arbitrary points are selected in the l-dimension space as the centers of the clusters. After that, all the documents are assigned to the nearest cluster center to minimize the sum of square error (SSE). Then, the center of each cluster can be updated by recalculating the mean of the documents in it. Finally, the above steps can be repeated until the centers do not change.

To select the value of k, the SSE performs as a key indicator by trying different values and plotting the SSE in a line chart. The value of k can be selected when the declining rate of SSE starts to become slow. Another measurement to determine the k value is called the "silhouette coefficient". It can show how similar one point is to the other points in the same cluster compared with the points of another cluster (de Barros Franco & Steiner, 2018). It can be expressed as:

$$s_i = \frac{b_i - a_i}{\max(a_i, b_i)} \quad (2)$$

where a_i is the average distance between document i and the other documents in the same cluster; and b_i is the average distance between document i and the ones in other clusters.

As the dimension of documents' contribution data may be larger than three, it will cause barriers to the visualization of the classification results. Therefore, PCA was applied to reduce the dimension to 2 or 3 before plotting the clustering of documents.

4. Results and discussion

Firstly, a word cloud based on the frequency of each keyword has been generated as shown in **Figure 2**. However, sufficient patterns would be found by separating the main themes of the whole collection of documents. Therefore, the bibliometric analysis was presented in section 4.1, following by the thematic analysis result in section 4.2, while the classification result of punishments is presented in section 4.3. Besides, the number of publications by year was shown in **Figure 3**. It shows the research area of sustainable UPM experiences a rapid development period during the past 15 years.

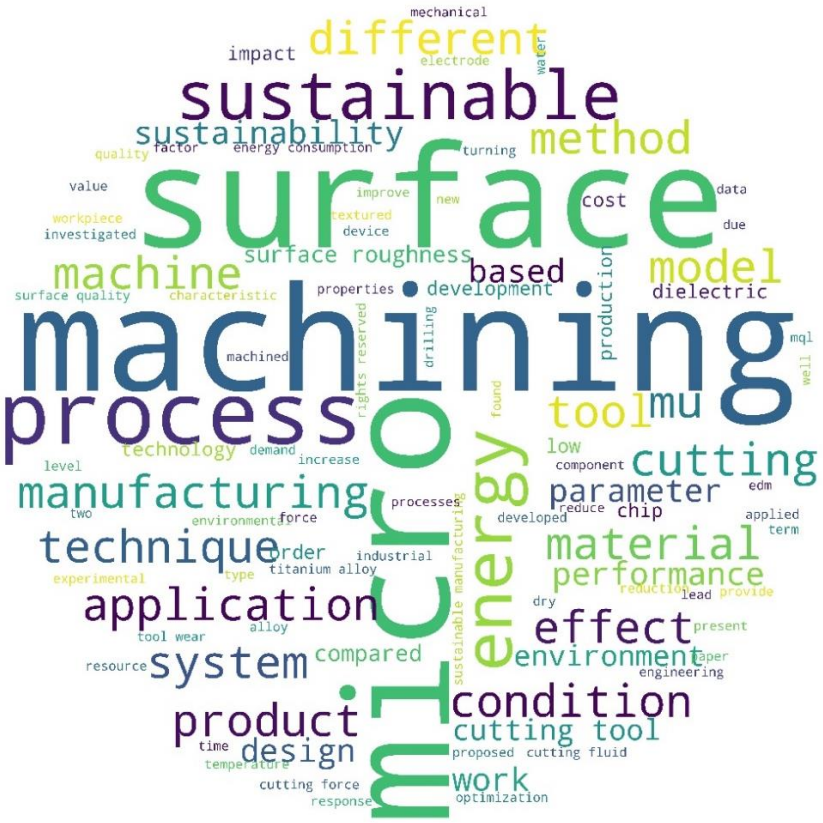

Figure 2 The word cloud of the sustainable UPM literature

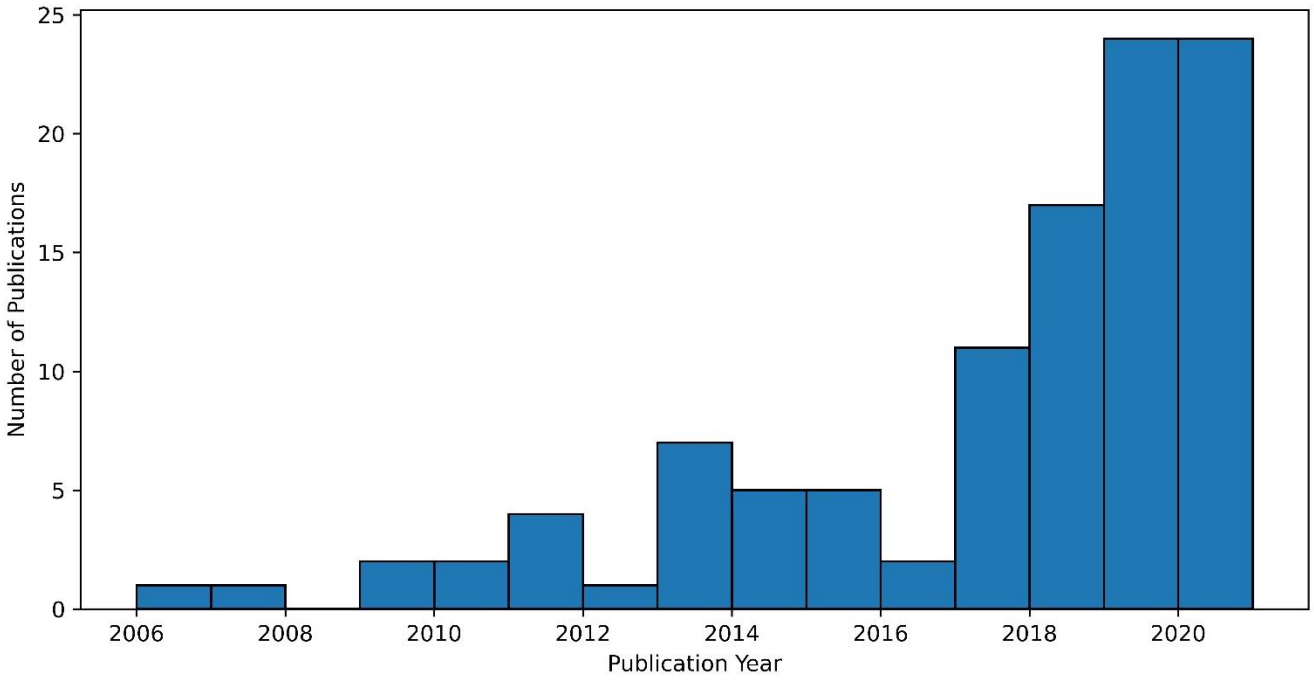

Figure 3 The distribution of sustainable UPM literature by year

4.1. Bibliometric analysis

Based on the bibliometric analysis of the selected publications by using VOSviewer, the citation network visualization of authors, sources, and organizations have been generated and presented in **Figure 4**, **Figure 5**, and **Figure 6** respectively. In these three figures, all the nodes without adjacent edges have been deleted as they cannot form a network. In **Figure 4**, different colors stand for clusters formed by authors according to their citation relationship. In this citation network, several connected components can be found (eg. The green cluster). And no or limited citations existed between one cluster with others can be found. It may be because the authors in the same cluster focused on a similar topic, more citing relationships that exist inside of the clusters. In the source citation network in **Figure 5**, the size of the node stands for the number of publications contributed by the source. It can be found the journal of cleaner production and the international journal of advanced manufacturing technology contribute more research articles on sustainable UPM than others. For the citation network of organizations, decentralized distribution of the nodes' size was found in **Figure 6**. It illustrates that the research area of sustainable UPM is not dominated by one or two institutions. However, most of the cross-institution citations exist in the largest connected component formed by the Hong Kong Polytechnic University, Brunel University London, and several other institutions in **Figure 6**.

Besides, the co-authorship network of authors, organizations, and countries has been visualized in **Figure 7**, **Figure 8**, and **Figure 9**. All the nodes connecting to no edge have been deleted as they cannot form a network. The network in **Figure 7** shows a decentralized distribution, which includes multiple dense connected components. As each connected component stands for a group of researchers who collaborate closely, this result illustrates the research area is not dominated by several research groups. Similarly, decentralized distribution was also found in the co-authorship network of organizations in **Figure 8**. **Figure 9** shows most works were contributed by China, the USA, and India as the size of node stands for the number of publications. And most cross-country collaboration happened among these three countries and several other countries, which formed the largest connected component together in **Figure 9**.

What's more, the top countries and authors of citation have also been ranked based on citation, which were shown in **Table 1**, and **Table 2**. It was shown that India and USA are the two countries with the most citations in this research field. What's more, South Korea is ranked at the third place of citation with only 6 publications, which reflected the publications from South Korea have a higher average impact. **Table 2** shows that Ahn, Sung-hoon, Yoon, Hea-sung, and Pham, Minh-quan were the top three authors with the highest citations, which all have 102 citations. It was because they were the co-authors in the same articles according to the raw data in the **Supplemental Material**.

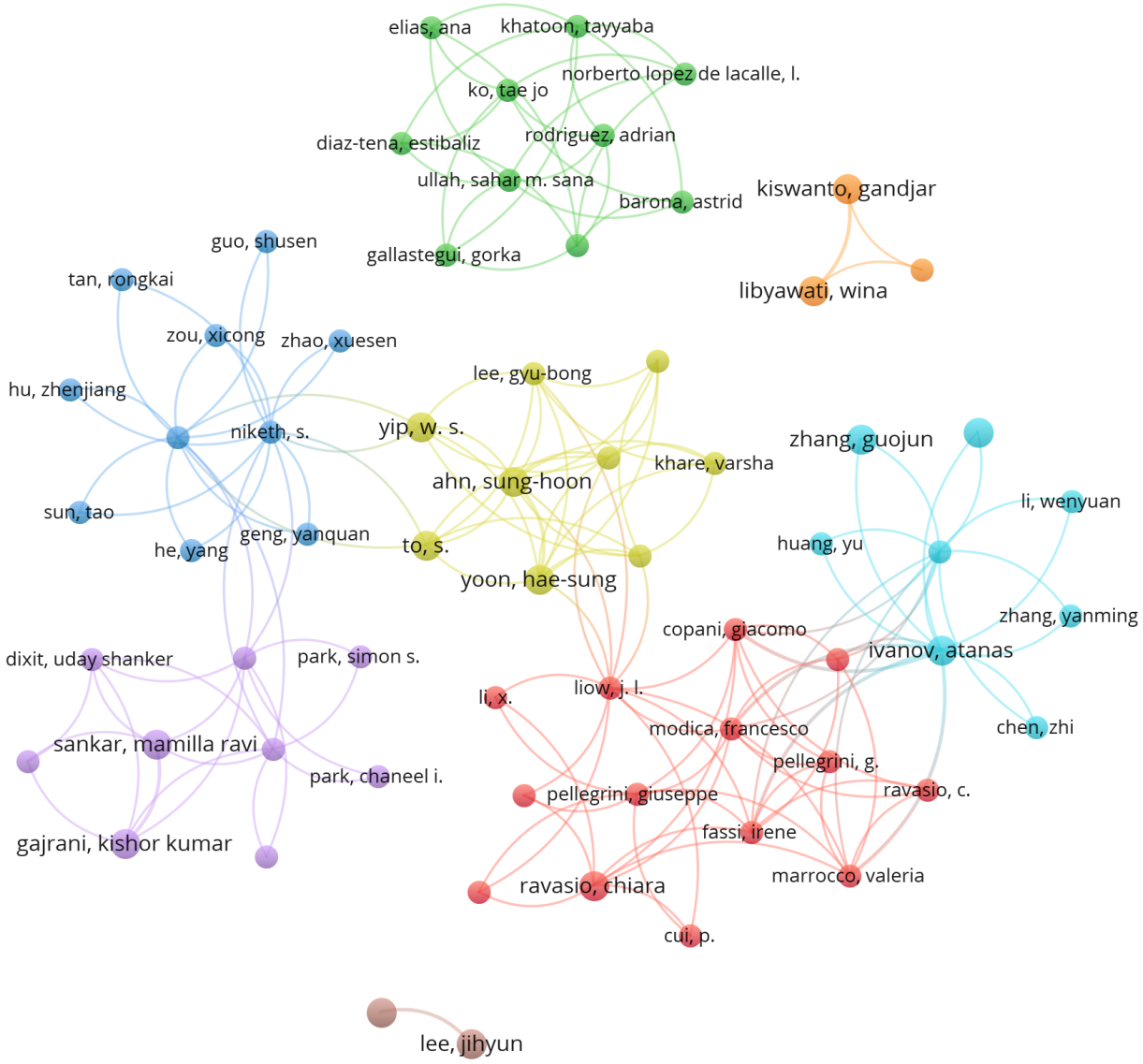

Figure 4 The authors' citation network in the research field of the sustainable UPM.

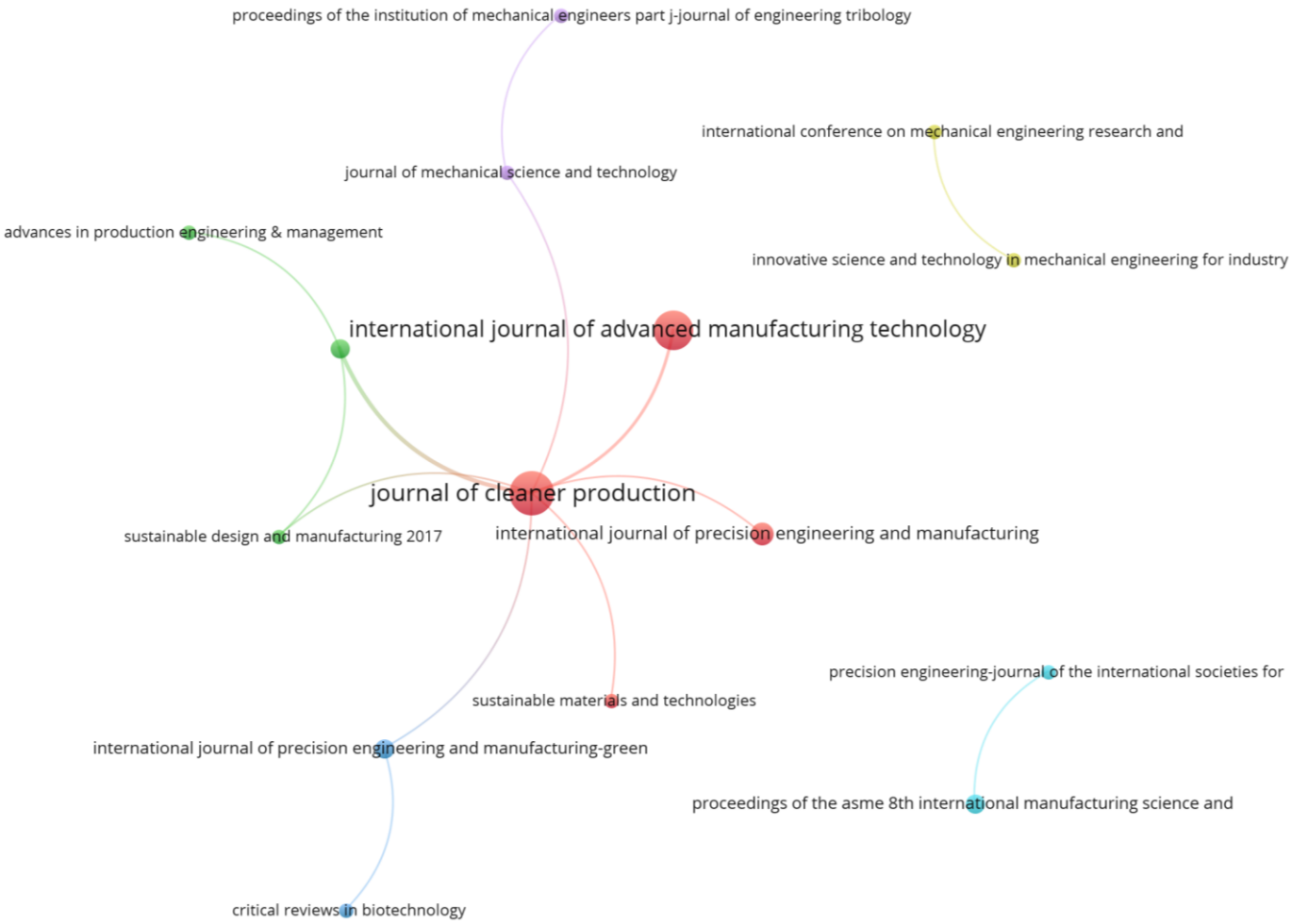

Figure 5 The sources' citation network in the research field of the sustainable UPM.

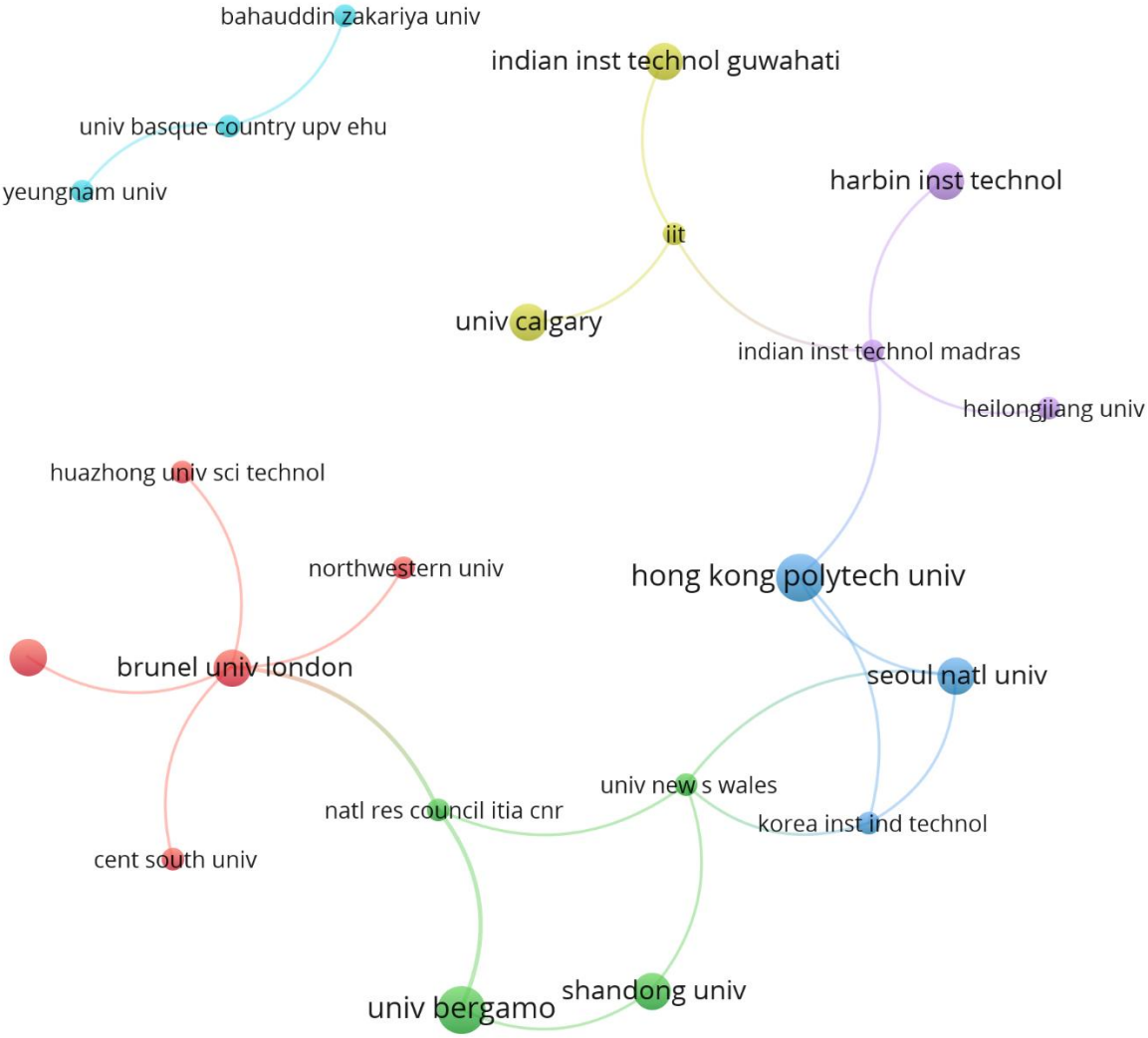

Figure 6 The organizations' citation network in the research field of the sustainable UPM.

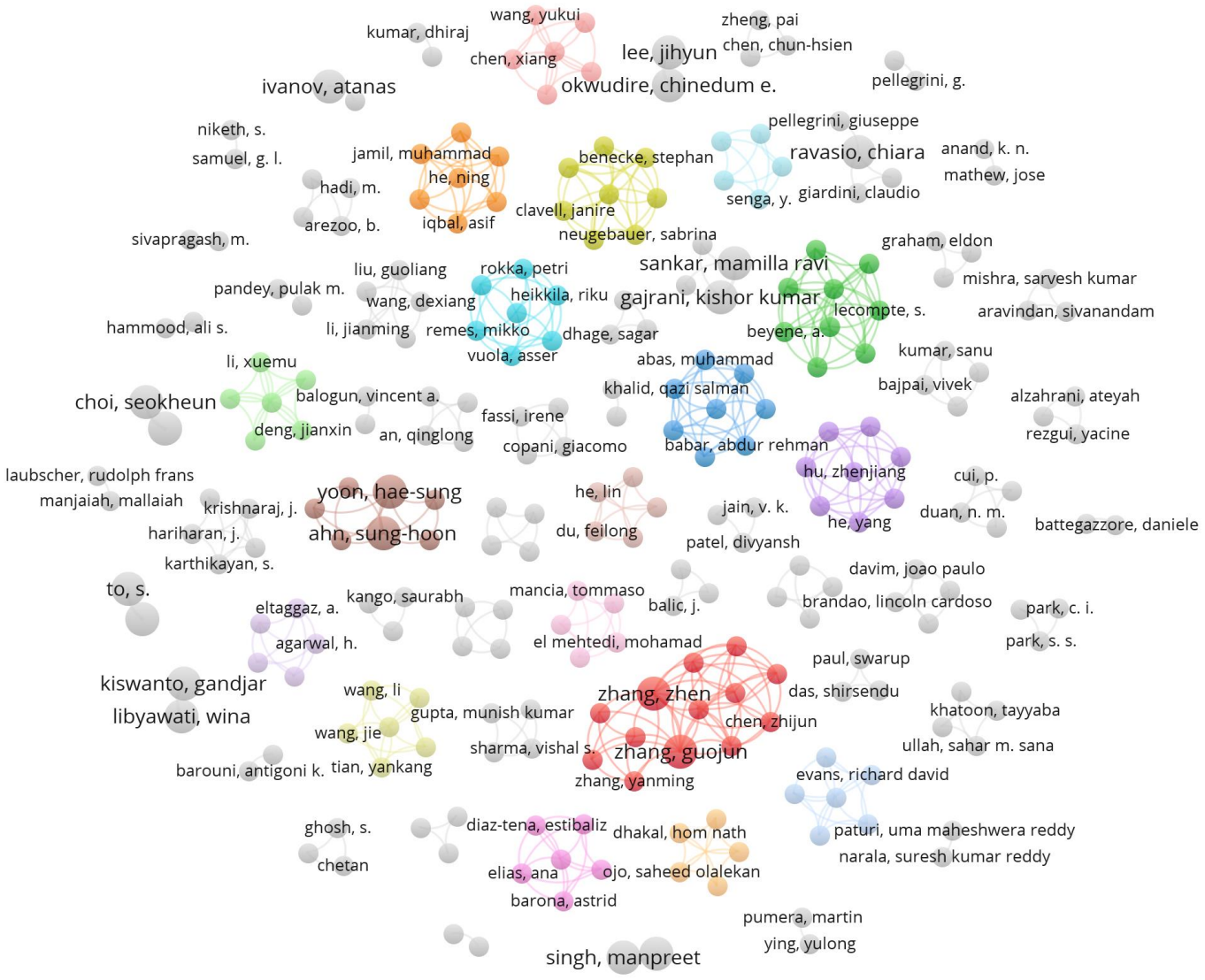

Figure 7 The co-authorship network of authors in the research field of the sustainable UPM.

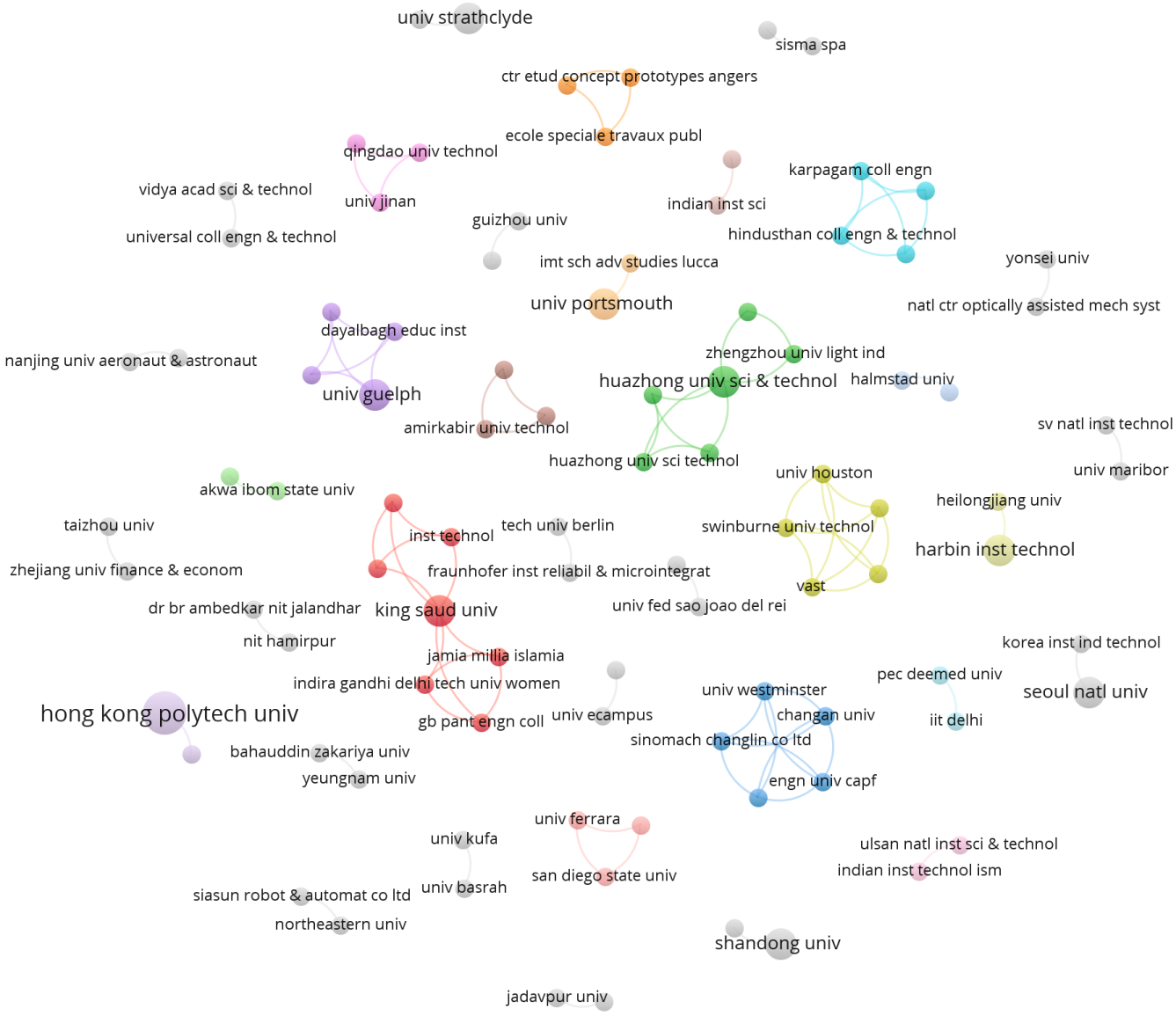

Figure 8 The co-authorship network of organizations in the research field of the sustainable UPM.

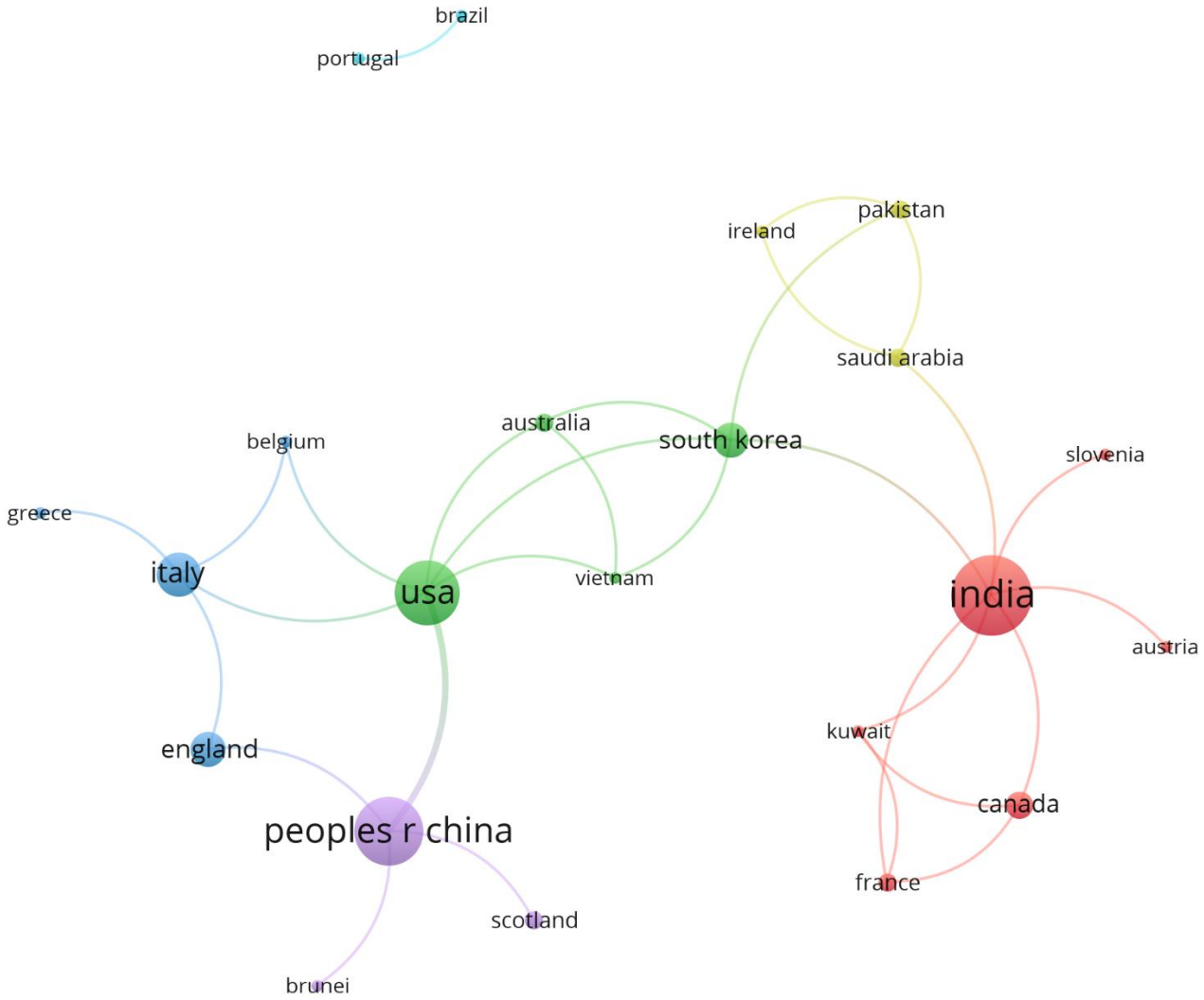

Figure 9 The co-authorship network of countries in the research field of the sustainable UPM.

Table 1 Rank of countries based on citation number

Rank	Country	Publications	Citations
1	India	26	353
2	USA	18	232
3	South Korea	6	110
4	China	20	107
5	Italy	9	78
6	Australia	2	54
7	England	6	51

8	Canada	4	45
9	Germany	3	40
10	France	2	37

Table 2 Rank of authors based on citation number

Rank	**Author**	**Publications**	**Citations**
1	ahn, sung-hoon	2	102
1	yoon, hea-sung	2	102
1	pham, minh-quan	2	102
4	pandey, pulak m.	1	57
4	sharma, varun	1	57
6	lee, gyu-bong	1	55
6	moon, jong-seol	1	55
8	liow,j.l.	1	53
8	kumar, arkadeep	1	53
8	melkote, shreyes n.	1	53

4.2. Thematic analysis results

The coherence score is a measurement used to evaluate topic modeling results based on the degree of semantic similarity of the words in one theme (Mohammed & Al-augby, 2020). And a large value of this metric indicates a high level of coherence. By utilizing the LDA model to evaluate the literature of sustainable UPM, the optimal number of topics is 4 according to the coherence score (as shown in **Figure 10**). Therefore, four main themes were identified and the top 10 keywords of each were presented in **Table 3**. The first theme includes the keywords like “machine”, “surface”, and “process”, which are related to the machining process. And the keywords of the second theme include “energy”, “sustainable”, “infrastructure” which focus on the application in sustainable energy. Besides that, the third theme consists of several keywords of technology details like “ air”, “spray”, “mwf”, and “pressure”. Therefore, this theme was named as machining technology. Lastly, the top three keywords of the fourth theme are “cleaning”, “surface”, and “self”, which indicates it is related to the research of surface structure with special patterns and functions like self-cleaning.

As the UPM is one cutting-edge technology with a shorter developing period compared with the traditional manufacturing method (Roeder, Guenther, & Zimmermann, 2019), the amount of sustainable UPM research

works are quite limited. Therefore, it is still in the emerging or growth stage in the technology life cycle with a high potential to be investigated. From the view of the TBL, the keywords of the main themes show the current research of sustainable UPM still concentrates on the technology aspect, while the economic aspect has not been discussed frequently. For the environment aspect, energy consumption and conservation are the most focused area. And the social aspect of sustainable UPM, like working conditions and product responsibility, still lacks investigation as no theme of sustainable UPM study includes this concept. As a comparison, the social sustainability performance of manufacturing and supply chain has been evaluated in various studies (Mani, Gunasekaran, & Delgado, 2018; Rajak & Vinodh, 2015). It indicates the researchers of sustainable UPM are from the field of engineering and science. It needs more economists and sociologists to analyze the economic and social aspects of UPM.

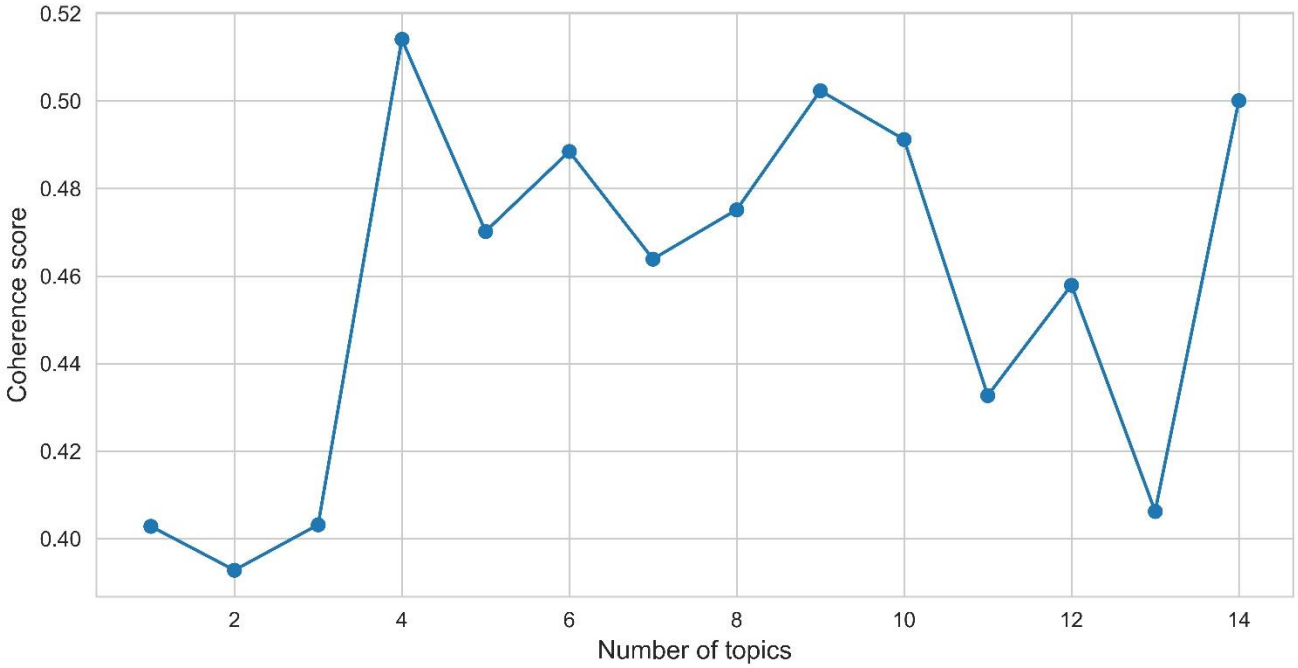

Figure 10 The coherence score for different numbers of themes

Table 3 The top 10 keywords for each theme

No.	Theme 1: machining process	Theme 2: sustainable energy	Theme 3: machining technology	Theme 4: surface structure
1	process	energy	technology	cleaning
2	machine	development	air	surface
3	surface	sustainable	rate	self
4	cut	construction	flow	production
5	impact	underground	tool	industrial
6	material	infrastructure	spray	cycle

7	tool	machine	machining	resource
8	machining	device	mwf (Metalworking fluids)	life
9	sustainable	technology	pressure	clean
10	develop	micro	wear	benefit

4.3. Classification of literature

As the SSE can measure the total internal distance of points in the same clusters, the object is to minimize the SSE value without increasing the number of clusters (the *k* value) too much. By plotting the line chart of SSE about *k*, the value of *k* can be considered the optimal one when the declining rate of the SSE starts to become negligible. Similarly, an optimal *k* value can get a sufficiently large Silhouette coefficient as the Silhouette coefficient is used to measure the external distance among the clusters. The SSE and Silhouette coefficient results were calculated and plotted in **Figure 11**. It shows that the declining rate of SSE started to slow down at the point of four. While the silhouette coefficient at this point is also considerably large. Therefore, the documents were classified into four groups according to the documents' contribution data by running the k-means algorithm. After that, PCA was utilized to reduce the dimension to 2 and 3 from 4 in order to polt the classification results in 2D and 3D figures. The eigenvalues and explained variation ratios of the first three principal components are shown in **Table 4**. These results show the first three principal components can already explain 99.5% of original data, which means the information loss was negligible. In this way, the documents before and after the classification are visualized in two dimensions and three dimensions (as shown by **Figure 12**). The detailed results of clustering and contribution percentage for each document are summarized in **Table *5***.

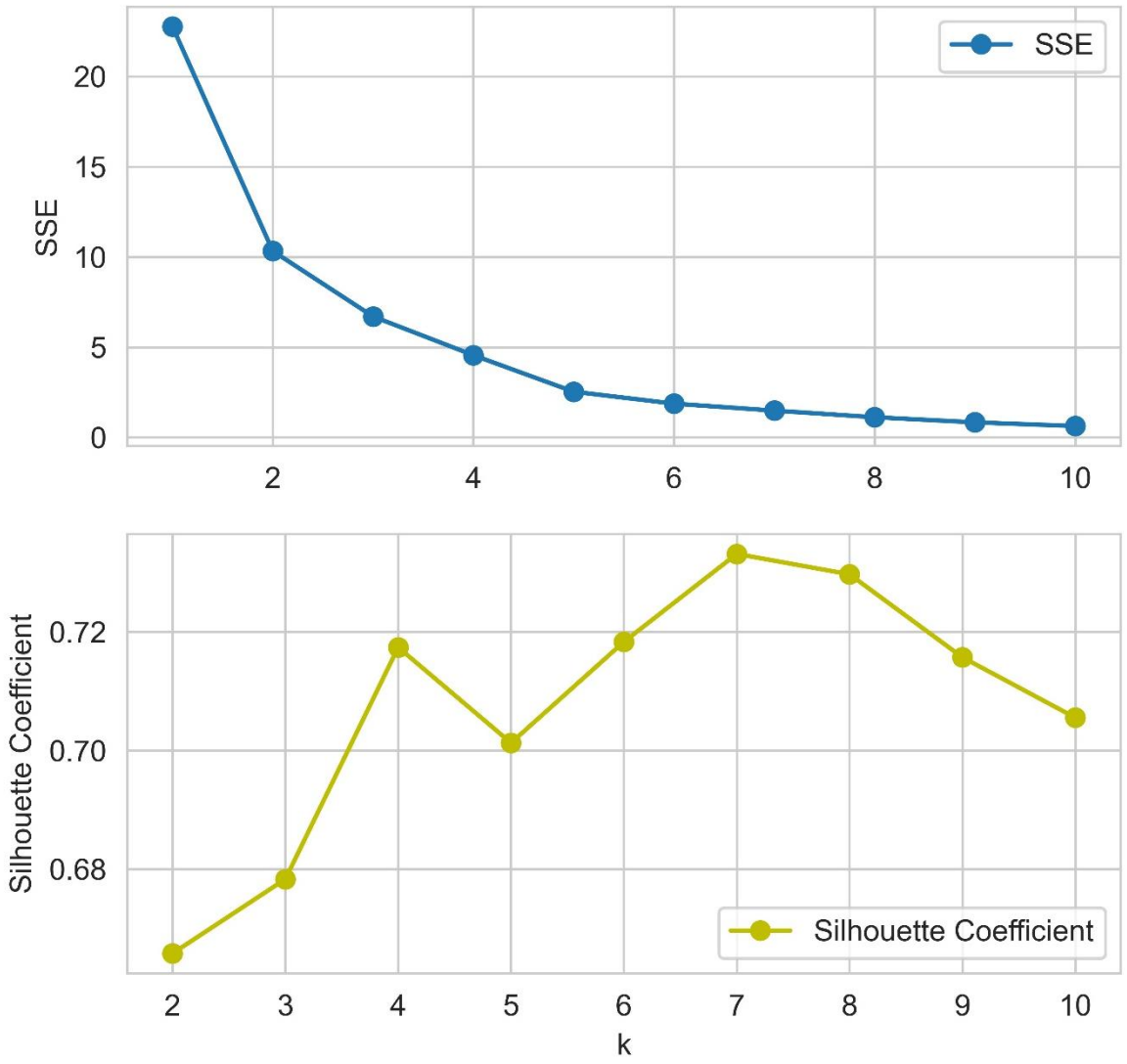

Figure 11 The SSE and Silhouette coefficient results

Table 4 The eigenvalues and explained variation ratios of the first three principal components

Principal components	Eigenvalue	Explained variation (%)
First	0.153	70.5%
Second	0.040	18.3%
Third	0.024	11.1%
Total	-	99.5%

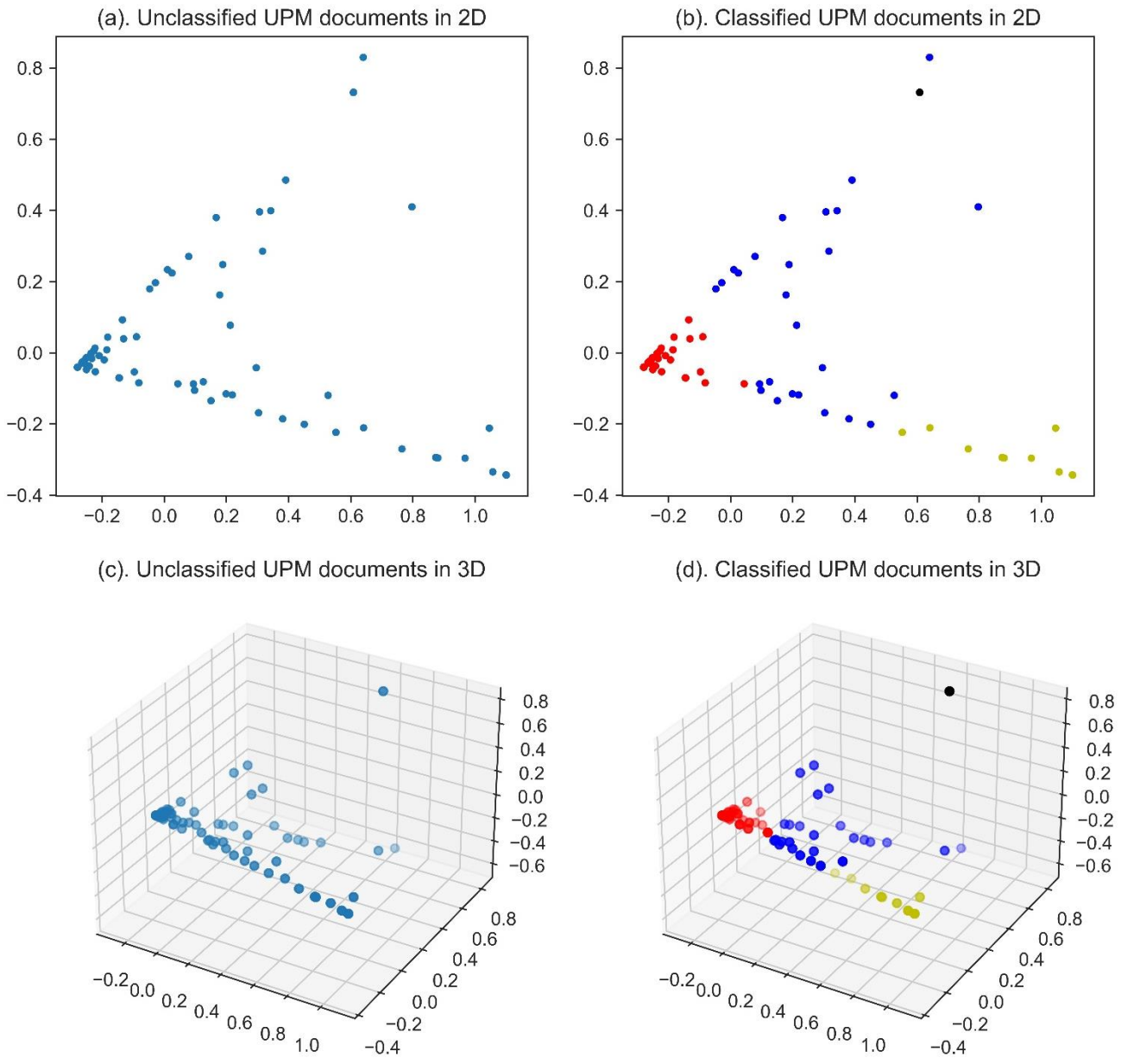

Figure 12 The 2D and 3D documents map of sustainable UPM research (documents in clusters 1, 2, 3, 4 are colored in red, yellow, blue, and black respectively)

From **Table 5**, it is shown that cluster 1 consists of 68 documents out of the total number. And for all the documents in this cluster, more than 75% of each document belongs to theme 1, which means the literature in cluster 1 focuses on theme 1. For cluster 2, it is formed by 10 documents and over 60% of each work contributes to theme 2. Moreover, a total of 26 documents belongs to cluster 3, and each of them contributes to multiple topics. Finally, cluster 4 only has 2 documents, which 100% belong to theme 3. Therefore, it can be found that clusters 1, 2, and 4 are distributed in the middle left, bottom right, and upper right corners respectively, while cluster 3 is distributed in the area among the other three clusters in **Figure 12** (d). It indicates that the works in cluster 4 have low similarity with the works in other clusters, which can provide more inspiration about widening the sustainable UPM fields.

Several key findings have been discovered based on the classification and the topics percentage in **Table 5**. Firstly, it indicates that the machining process is the hottest topic in the domain of sustainable UPM as more than 60% of works focus on it. The top three keywords of topic 1 are "process", "machine", and "surface", which shows the most focused components in this topic. The second topic was covered by around 10% of publications in cluster 2, which makes it the second hot topic in this area. What's more, theme 3 consists of keywords like "air", "flow", "spray" and "dry", which are related to technical details of UPM. However, this theme was only focused by only 2 papers, which means it is still in the emerging stage. For the last theme, only 2 papers contribute more than 60% of content to it, which cannot form an individual cluster. However, 15 publications contribute 10%-60% of the content in this theme, which illustrates theme 4 is normally studied with other themes. For the researchers interested in the interdisciplinary study of sustainable UPM, theme 4 could provide more opportunities. Besides theme 4, most of the publications in cluster 4 also belong to interdisciplinary studies among multiple themes. This cluster may present more creative ideas by trying different methods to investigate sustainable UPM.

Table 5 The clustering and contribution percentage of sustainable UPM literature

Cluster	**Publication Title**	**Topic 1 %**	**Topic 2 %**	**Topic 3 %**	**Topic 4 %**
1	46th ESAO Congress 3-7 September 2019 Hannover, Germany Abstracts	96.8%	2.0%	0.0%	1.1%
	A new sustainable direct solid state recycling of AA1090 aluminum alloy chips by means of friction stir back extrusion process	99.9%	0.0%	0.0%	0.0%
	A REVIEW OF THE SURFACE MODIFICATIONS OF TITANIUM ALLOYS FOR BIOMEDICAL APPLICATIONS	90.2%	9.8%	0.0%	0.0%
	A Study on New Surface Textured Tools for Inhibition of Derivative Cutting	99.8%	0.0%	0.0%	0.0%
	A sustainability index for the micro-EDM drilling process	97.5%	1.3%	1.2%	0.0%
	A Sustainable solid state recycling of pure Aluminum by means of Friction Stir Extrusion process (FSE)	99.8%	0.0%	0.0%	0.0%
	Analysis and reduction of process energy consumption and thermal deformation in a micro-structure wire electrode electric discharge machining thin-wall component	95.3%	1.1%	3.6%	0.0%
	Analyzing sustainable performance on high-precision molding process of 3D ultra-thin glass for smart phone	75.6%	21.8%	2.6%	0.0%
	Atomized Dielectric Spray-Based Electric Discharge Machining for Sustainable Manufacturing	89.0%	0.0%	11.0%	0.0%
	Behaviour of a biocompatible titanium alloy during orthogonal micro-cutting employing green machining techniques	99.8%	0.0%	0.0%	0.0%
	Biomachining: metal etching via microorganisms	94.1%	0.0%	4.6%	1.2%
	Chip formation and hole quality in dry drilling additive manufactured Ti6Al4V	97.9%	0.0%	1.2%	0.0%
	Combustion Simulation and Emission Prediction of Different Combustion Chamber Geometries Using Finite Element Method	85.7%	11.5%	2.8%	0.0%
	Comparison between sustainable cryogenic techniques and nano-MQL cooling mode in turning of nickel-based alloy	99.9%	0.0%	0.0%	0.0%

	Comparison of a Single-Screw and a Scroll Expander under Part-Load Conditions for Low-Grade Heat Recovery ORC Systems	99.9%	0.0%	0.0%	0.0%
	Damage investigation and assessment due to low-velocity impact on flax/glass hybrid composite plates	99.8%	0.0%	0.0%	0.0%
	Design and Analysis for Vibration Assisted Micro Milling	97.8%	2.1%	0.0%	0.0%
	Development of Empirical Model for Biomachining to Improve Machinability and Surface Roughness of Polycrystalline Copper	93.2%	2.0%	3.4%	1.4%
	Development of Pneumatic Projector for Environmentally Sustainable Impact Based Nanomachining Operations and Nanoparticle Deposition	99.8%	0.0%	0.0%	0.0%
	Diamond Wire Sawing of Solar Silicon Wafers: A Sustainable Manufacturing Alternative to Loose Abrasive Slurry Sawing	99.9%	0.0%	0.0%	0.0%
	Effects of Electrode and Workpiece Materials on the Sustainability of Micro-EDM Drilling Process	99.9%	0.0%	0.0%	0.0%
	Effects of surface texture parameters of cutting tools on friction conditions at tool-chip interface during dry machining of AISI 1045 steel	97.5%	0.0%	2.4%	0.0%
	Effects of the influence factors in adhesive workpiece clamping with ice: experimental study and performance evaluation for industrial manufacturing applications	95.8%	4.2%	0.0%	0.0%
	Enhancing the tribological aspects of machining operation by hybrid lubrication-assisted side-flank-face laser-textured milling insert	96.8%	0.0%	0.0%	3.1%
	Environmentally friendly machining with MoS2-filled mechanically microtextured cutting tools	96.8%	0.0%	0.0%	3.1%
	Evaluation of ionic liquids as lubricants in micro milling - process capability and sustainability	99.9%	0.0%	0.0%	0.0%
	Evaluation of power consumption and MRR in WEDM of Ti-6Al-4V alloy and its simultaneous optimization for sustainable production	99.9%	0.0%	0.0%	0.0%
	Evaluation of size effect and improvement in surface characteristics using sunflower oil-based MQL for sustainable micro-endmilling of Inconel 718	99.9%	0.0%	0.0%	0.0%
	Evaluation of specific cutting energy considering effects of cutting tool geometry during micro-machining process	99.9%	0.0%	0.0%	0.0%
	Evaluation of the sustainability of the micro-electrical discharge milling process	99.9%	0.0%	0.0%	0.0%
	Experimental Investigation and Statistical Evaluation of Optimized Cutting Process Parameters and Cutting Conditions to Minimize Cutting Forces and Shape Deviations in Al6026-T9	99.9%	0.0%	0.0%	0.0%
	Experimental Investigation on Machinability of Aluminum Alloy during Dry Micro Cutting Process Using Helical Micro End Mills with Micro Textures	97.0%	0.0%	3.0%	0.0%
	Experimental investigation to study the effect of electrostatic micro-solid lubricant-coated carbide tools on machinability parameters in turning	99.9%	0.0%	0.0%	0.0%
	FABRICATION OF CHITOSAN POROUS STRUCTURE AND APPLICATIONS ON ARTIFICIAL PHOTOSYNTHESIS DEVICE	99.9%	0.0%	0.0%	0.0%
	Fundamental Aspects in Designing Vibration Assisted Machining: A Review	99.9%	0.0%	0.0%	0.0%
	Investigation of Induction Motor Support in weak Microgrids using Virtual Synchronous Generator	98.4%	0.0%	0.0%	1.5%
	Investigation of surface integrity in high speed milling of gamma titanium aluminide under dry and minimum quantity lubricant conditions	99.8%	0.0%	0.0%	0.0%
	Machinability investigation and sustainability analysis of minimum quantity lubrication-assisted micro-milling process	99.9%	0.0%	0.0%	0.0%

	Machining damage and surface integrity evaluation during milling of UD-CFRP laminates: vs.	98.3%	0.0%	0.0%	1.6%
	Machining performance evaluation of Ti6Al4V alloy with laser textured tools under MQL and nano-MQL environments	94.9%	0.0%	5.0%	0.0%
	Micro texturing on metallic surfaces: State of the art	90.9%	2.5%	0.0%	6.5%
	Micro/Nanomotors for Water Purification	84.0%	0.0%	5.7%	10.2%
	Minimization of the residual vibrations of ultra-precision manufacturing machines via optimal placement of vibration isolators	99.9%	0.0%	0.0%	0.0%
	Modelling and Analysis of the Temperature Distribution of a Micro-channel Internally Cooled Smart Cutting Tool in Machining AlSi7	99.8%	0.0%	0.0%	0.0%
	Multi-objective optimization of machining and micro-machining processes using non-dominated sorting teaching-learning-based optimization algorithm	99.9%	0.0%	0.0%	0.0%
	Multi-response optimisation of micro-milling performance while machining a novel magnesium alloy and its alumina composites	99.8%	0.0%	0.0%	0.0%
	Multi-response optimization for Nimonic alloy miniature gear fabrication using wire electrical discharge machining	99.9%	0.0%	0.0%	0.0%
	OPTIMAL MOTOR LOCATION FOR THE REDUCTION OF RESIDUAL VIBRATIONS IN MODE-COUPLED ULTRA-PRECISION MANUFACTURING MACHINES	99.9%	0.0%	0.0%	0.0%
	Optimization and fabrication of curvilinear micro-grooved cutting tools for sustainable machining based on finite element modelling of the cutting process	99.9%	0.0%	0.0%	0.0%
	Optimization of titanium alloys turning operation in varied cutting fluid conditions with multiple machining performance characteristics	99.9%	0.0%	0.0%	0.0%
	Performance Optimization of a Kirsten-Boeing Turbine by A Metamodel Based on Neural Networks Coupled with CFD	96.9%	0.0%	0.0%	3.0%
	Prediction of Surface Roughness of 304 Stainless Steel and Multi-Objective Optimization of Cutting Parameters Based on GA-GBRT	99.9%	0.0%	0.0%	0.0%
	Recent advances in turning with textured cutting tools: A review	99.9%	0.0%	0.0%	0.0%
	Replacement of Hazard Lubricants by Green Coolant in Machining of Ti6Al4V: A 3D FEM Approach	99.9%	0.0%	0.0%	0.0%
	Sensing, smart and sustainable product development (S-3 product) reference framework	85.6%	14.3%	0.0%	0.0%
	Surface texturing for tribology enhancement and its application on drill tool for the sustainable machining of titanium alloy	93.8%	0.0%	0.0%	6.2%
	Surface topography characterization of brass alloys: lead brass (CuZn39Pb3) and lead free brass (CuZn21Si3P)	99.9%	0.0%	0.0%	0.0%
	Sustainability analyses of cutting edge radius on specific cutting energy and surface finish in side milling processes	99.9%	0.0%	0.0%	0.0%
	Sustainable machining of aerospace material - Ti (grade-2) alloy: Modeling and optimization	99.9%	0.0%	0.0%	0.0%
	Sustainable Manufacturing of Precision Miniature Gears by Abrasive Water Jet Machining-An Experimental Study	85.7%	3.8%	0.0%	10.5%
	Sustainable manufacturing of ultra-precision machining of titanium alloys using a magnetic field and its sustainability assessment	90.3%	9.7%	0.0%	0.0%
	Sustainable Micro-Manufacturing of Micro-Components via Micro Electrical Discharge Machining	92.3%	3.6%	4.1%	0.0%
	Sustainable Milling of Ti-6Al-4V: Investigating the Effects of Milling Orientation, Cutter's Helix Angle, and Type of Cryogenic Coolant	99.9%	0.0%	0.0%	0.0%
	Sustainable mu ECM machining process: indicators and assessment	82.2%	5.9%	0.0%	11.9%

	Sustainable production of dry-ultra-precision machining of Ti-6Al-4V alloy using PCD tool under ultrasonic elliptical vibration-assisted cutting	95.1%	0.0%	3.4%	1.5%
	Sustainable Ultra-Precision Machining of Titanium Alloy Using Intermittent Cutting	99.9%	0.0%	0.0%	0.0%
	Toward sustainable micro-drilling of Inconel 718 superalloy using MQL-Nanofluid	99.9%	0.0%	0.0%	0.0%
	Tribo-mechanical and surface morphological comparison of untextured, mechanical micro-textured (M mu T), and coated-M mu T cutting tools during machining	99.9%	0.0%	0.0%	0.0%
2	A micro-sized bio-solar cell for self-sustaining power generation	0.0%	88.3%	5.1%	6.6%
	A Micro-sized Microbial Solar Cell	9.1%	89.1%	0.0%	1.8%
	A Novel Microalgae Harvesting Method Using Laser Micromachined Glass Fiber Reinforced Polymers	32.3%	64.7%	0.0%	3.0%
	Biomass based bio-electro fuel cells based on carbon electrodes: an alternative source of renewable energy	24.4%	75.6%	0.0%	0.0%
	Johns Hopkins on the chip: microsystems and cognitive machines for sustainable, affordable, personalised medicine and healthcare	16.0%	83.9%	0.0%	0.0%
	Mechanical micromachining: a sustainable micro-device manufacturing approach?	0.0%	99.7%	0.0%	0.0%
	MICRO AND NANO DESIGN AND FABRICATION OF A NOVEL ARTIFICIAL PHOTOSYNTHESIS DEVICE	16.5%	83.5%	0.0%	0.0%
	Modelling and implementing smart micro-grids for fish-processing industry	3.2%	96.7%	0.0%	0.0%
	PARAMETRIC 4D MODELING FOR REAL-TIME VISUALIZATION OF MICROTUNNELING OPERATIONS	0.0%	99.9%	0.0%	0.0%
	Study of the skid resistance of blends of coarse aggregates with different polish resistances	39.7%	60.2%	0.0%	0.0%
3	Abrasive water jet drilling of advanced sustainable bio-fibre-reinforced polymer/hybrid composites: a comprehensive analysis of machining-induced damage responses	65.7%	0.0%	34.3%	0.0%
	Addressing sustainability and flexibility in manufacturing via smart modular machine tool frames to support sustainable value creation	55.8%	13.2%	27.5%	3.6%
	Bio-based PA5.10 for Industrial Applications: Improvement of Barrier and Thermo-mechanical Properties with Rice Husk Ash and Nanoclay	68.4%	0.0%	0.0%	31.5%
	Control of machining parameters for energy and cost savings in micro-scale drilling of PCBs	69.1%	26.3%	0.0%	4.5%
	Electrochemical discharge machining: fumes generations, properties and biological effects	47.1%	52.8%	0.0%	0.0%
	Enhanced carbide tool life by the electromagnetic coupling field for sustainable manufacturing	59.6%	0.0%	40.3%	0.0%
	Fabrication of Micro-Dimpled Surfaces through Micro Ball End Milling	72.1%	26.2%	0.0%	1.7%
	Feasibility assessment of some alternative dielectric mediums for sustainable electrical discharge machining: a review work	51.3%	0.0%	3.9%	44.8%
	Force modeling and applications of inclined ball end milling of micro-dimpled surfaces	62.9%	34.1%	0.0%	3.0%
	Improved Field Emission Properties of Carbon Nanostructures by Laser Surface Engineering	52.1%	47.8%	0.0%	0.0%
	Industrial Clusters and Directions of Technological Innovation in Taizhou	52.3%	8.3%	30.3%	9.1%
	Industrial smart product-service systems solution design via hybrid concerns	37.0%	9.3%	0.0%	53.7%
	Integrated, multidisciplinary approaches for micro-manufacturing research, and new opportunities and challenges to micro-manufacturing	38.3%	52.1%	4.2%	5.4%
	LIFE CYCLE INVENTORY STUDY OF BIOLOGICALLY INSPIRED SELF-CLEANING SURFACES	0.0%	0.0%	0.0%	99.9%
	Manufacturing Process Innovation-Oriented Knowledge Evaluation Using MCDM and Fuzzy Linguistic Computing in an Open Innovation Environment	39.9%	7.4%	0.0%	52.6%
	Micro-texturing of a WC-10Co-4Cr-Coated ASTM A479 Steel to Form a Super-Hydrophobic Surface	30.5%	6.8%	0.0%	62.7%

	On Energy Efficient and Sustainable Machining through Hybrid Processes	57.1%	21.4%	4.7%	16.8%
	Progress of remanufacturing engineering and future technology expectation	0.0%	34.8%	12.0%	53.1%
	Self-Sustainable Robotic Environment Discovery for Energy Harvesting Internet of Things	71.8%	24.8%	0.0%	3.3%
	Single Roll Melt Spinning Technique Applied as a Sustainable Forming Process to Produce Very Thin Ribbons of 5052 and 5083 Al-Mg Alloys Directly from Liquid State	64.2%	32.6%	3.1%	0.0%
	Smart Control of a Future Micro-grid in Olst	74.7%	0.0%	0.0%	25.2%
	Solutions to Mitigate the Impact of Electrical Machines Environment	54.4%	33.9%	0.0%	11.7%
	Sustainability of Micro Electrochemical Machining: Discussion	68.8%	31.1%	0.0%	0.0%
	Sustainable electrochemical discharge machining process: Characterization of emission products and occupational risks to operator	57.8%	42.2%	0.0%	0.0%
	Sustainable Manufacturing of High-Precision, Heat-Resistant Aspherical Lenses Using Ultraviolet Illumination With Prognosis of Remaining Useful Life	42.4%	14.6%	3.4%	39.5%
	Sustainable production of micro gears combining micro reciprocated wire electrical discharge machining and precision forging	72.7%	0.0%	0.0%	27.3%
4	Micro-flood (MF) technology for sustainable manufacturing operations that are coolant less and occupationally friendly	0.0%	0.0%	99.9%	0.0%
	Miniaturization of Flexible Screwing Cell	0.0%	0.0%	99.8%	0.0%

5. Research and managerial implications

Based on the results in section 4, the below implications are suggested:

1. One of the main challenges in the sustainable UPM area is the lack of studies about UPM's social aspects. As pointed by Malek and Desai (2020), the food industry, steel industry, and chemical industry have high potentials to improve social sustainability as they may cause more social impacts. The ultra-precision products involving chemical processes need more evaluation on their potential influences on social sustainability by sociologists, especially the experts of computational social science.
2. It was found that the life cycle assessment (LCA) of UPM products is still scarce, which is commonly used in sustainable manufacturing research (Gbededo, Liyanage, & Garza-Reyes, 2018). Therefore, LCA could provide quite a few research opportunities to evaluate the environmental impacts of UPM products like ultra-precision lenses.
3. Material waste is another direction that can bring new room for investigation. Some widely used tools in the industrial-ecology discipline, such as material flow analysis (Islam & Huda, 2019) and multi-criteria decision-making methods (Torkayesh, Malmir, & Asadabadi, 2021), shows their potential to improve the efficiency of waste management.

6. Model comparison and validation

In this section, the LDA model used in this work was compared with a commonly used keywords classification method named VOS to illustrate the advantages of LDA in the thematic analysis. Besides, the LDA model was applied to analyze the sustainable manufacturing research documents in section 6.2 to validate the adaptability in a different dataset.

6.1. Comparison between LDA and VOS Models

To show the advantages of the clustering method based on LDA in this study, the results in **Table 3** were compared with one commonly used keywords clustering technique created by Van Eck and Waltman (2010), which is called as VOS method. It was pointed out that VOS is a more effective method than the multidimensional scaling (MDS) technique (Van Eck, Waltman, Dekker, & van den Berg, 2010). By using the VOSviewer, a co-occurrence network of keywords can be visualized as shown in **Figure 13**. In this network, the size of a node represents the occurrences of the keyword (Ding, Rousseau, & Wolfram, 2016), while the link stands for the number of co-occurrences of these two keywords.

The first difference between the two models is that the keyword in VOS can only be assigned in only one cluster (Van Eck & Waltman, 2013). The LDA method allows repeated keywords in different topics as it is a probabilistic model (Mohammed & Al-augby, 2020). The main idea of the keywords classification in LDA is that the words with similar meanings tend to occur in similar contexts (Aletras & Stevenson, 2013). Some words may appear with different sets of words with highly various meanings as one word could have different meanings in different scenarios. Besides, if one word is suitable for different themes among the selected documents, it also has a high probability to be included in multiple clusters (themes). For example, as all the selected publications are relevant to sustainable development and machining, the word “sustainable” and “ machine” were included in more than one theme. What’s more, the output of the LDA model is the probability distribution of each word in the documents that belong to each theme (Negara, Triadi, & Andryani, 2019). Only the top ten keywords with the highest probability that belongs to the theme were selected. One keyword may be ranked at the top ten keywords list for more than one theme. In this way, the LDA method can classify the keywords based on the degree of “appearing together”. For the VOS method, it may lead to missing information of clusters if each keyword can only be assigned in one cluster. This problem could bring difficulties in interpreting the theme of each cluster, especially when the document size is small. As shown in **Figure 13**, each of clusters 7 and 8 only includes three keywords. If more keywords in other clusters can also belong to clusters 7 and 8 at the same time, it will be easier to explain the main theme of clusters 7 and 8.

Apart from that, another advantage of the LDA method is that it enables the determination of the optimal number of themes (keywords clusters) by using the coherence score mentioned in section 3.3. While the VOS method only allows adjusting the number of clusters by changing the minimum cluster size. And there is no standard method to determine the value of minimum cluster size in the VOS model, which may cause an

inadequate number of keywords in one cluster. As shown in **Figure 13**, eight clusters were obtained by using the default setting of minimum cluster size. Cluster 7 consists of only three keywords: “microhardness”, “surface integrity”, and “minimum quantity lubrication”, which is difficult to summarize the theme or topic of this keywords cluster. Therefore, the LDA model is more suitable for conducting the thematic analysis of documents compared with the VOS method.

Moreover, the keywords in the VOS model could be formed by more than one word as they were labeled by the authors of the articles. While the keywords in the LDA model only contain one word, which was extracted from the abstracts by the algorithm. In the LDA algorithm, each word is tokenized and put into an array (Helal & Mouhoub, 2018). To generate the final output, every word has a corresponding probability for each theme (Negara et al., 2019), while all the probabilities of the words will sum up to one. Therefore, it is still difficult for the current LDA model to include two-word phrases or even three-word phrases.

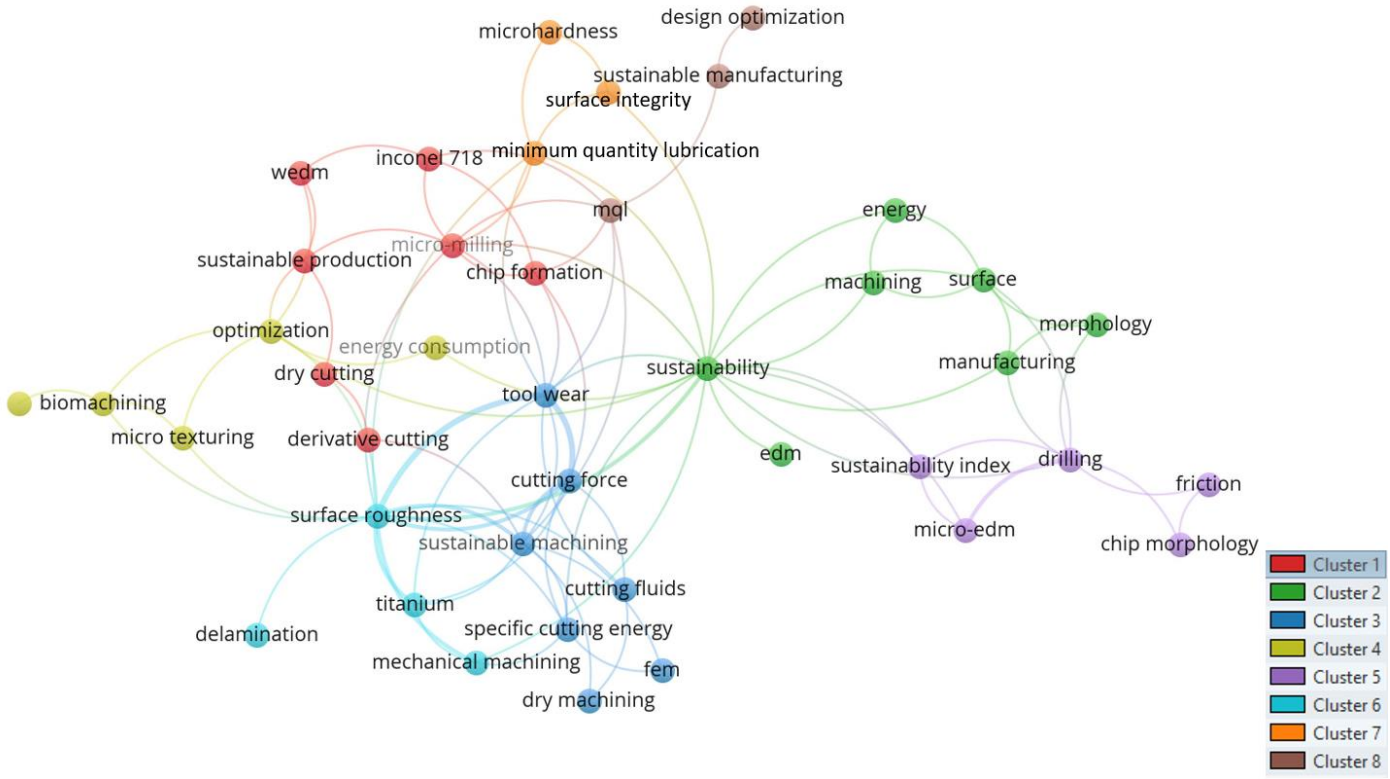

Figure 13 The classification of keywords based on co-occurrence of keywords

6.2. Validation of the model adaptability

As mentioned by Pang, Ban, Kadobayashi, and Kasabov (2011), one advantage of the LDA model is its adaptability in a wide range of non-linear problems. To illustrate the adaptability of the LDA model used in this work, it was applied to conduct the thematic analysis of the sustainable manufacturing (SM) field. In this section, a total of 6719 relevant research works were collected from Web of Science as raw data. The

coherence score for different numbers of themes is shown in **Figure 14**, which shows the optimal number of themes is 6 for this dataset. And the top 10 keywords of each theme are presented in **Table 6**. Based on the meanings of these keywords, the six themes were named as “production”, “environment”, “entrepreneur”, “entrepreneur”, “resource”, “economic”, and “waste treatment”. It can be found most keywords in one theme are relevant to one main topic, which enables the interpretation of the themes. For example, theme 3 contains the keywords: “company”, “firm”, “business”, which were highly relevant to entrepreneur. It shows the LDA model has reliable adaptability to analyze different types of textual datasets.

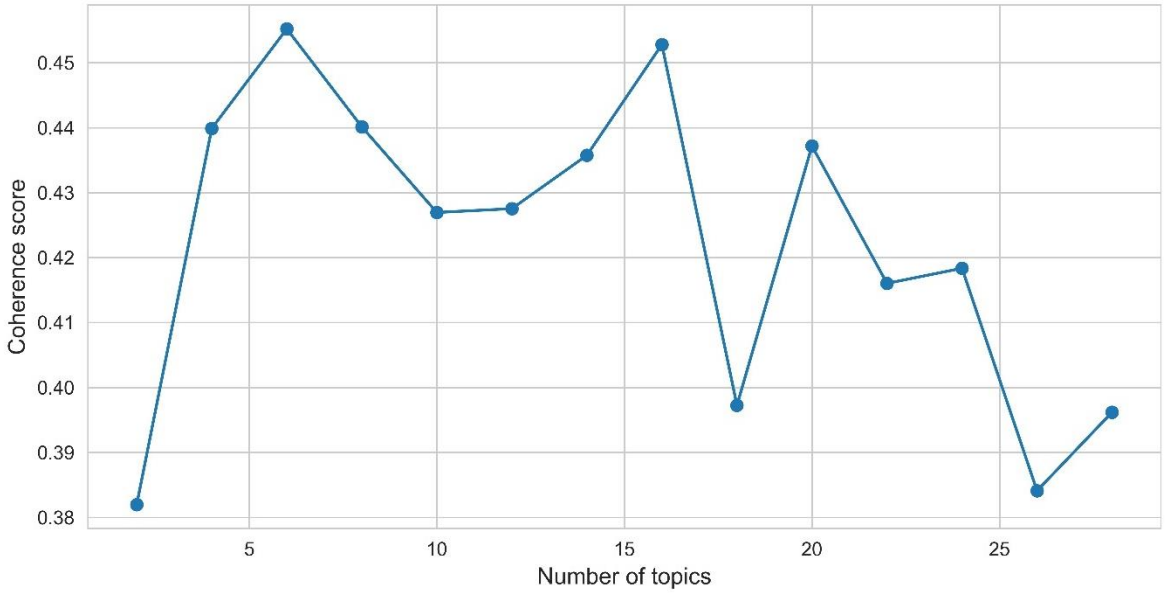

Figure 14 The Coherence scores for different number of themes in the sustainable manufacturing field

Table 6 The top 10 keywords for each theme of sustainable manufacturing research works

No.	Theme 1: production	Theme 2: environment	Theme 3: entrepreneur	Theme 4: resource	Theme 5: economic	Theme 6: waste treatment
1	product	energy	environmental	material	economic	chemical
2	sustainable	environmental	company	waste	development	high
3	manufacturing	impact	firm	concrete	growth	water
4	process	process	performance	manufacture	industry	treatment
5	design	cost	management	result	sector	forest
6	system	emission	practice	recycle	policy	process
7	paper	consumption	study	natural	market	production
8	technology	production	business	construction	resource	produce
9	manufacture	reduce	sustainable	study	country	use
10	development	result	manufacturing	property	economy	sustainable

7. Conclusion

In this paper, a thematic analysis model based on LDA and unsupervised learning algorithms was built up, which was utilized to find out different topics in the overall field of sustainable UPM. As a result, a total of four main themes were identified: machining process, sustainable energy, machining technology, and surface structure. After that, the content percentage of each document's contributing to these four themes can also be collected. By using the document's contribution percentage as raw data for the k-means algorithm, all the publications can be classified into 4 groups. This result illustrated that the machining process is the most focused theme in the current research about sustainable UPM. And most of the research about surface structure belongs to interdisciplinary study, which encourages innovation in methodology.

To be summarized, this work has four main contributions. Firstly, this work conducted a thematic analysis of the current research about sustainable UPM by using the text mining method. Compared with the work of Schneider et al. (2019), it delivers the details of the main sub-themes in this field as well as the classification of literature based on their contributions to each theme. In the methodology aspect, it developed a novel thematic analysis method by combining LDA and unsupervised learning methods for the first time. What's more, the technology life cycle of sustainable UPM has also been discussed for the first time. Additionally, several implementations have been given to the researchers who are interested in this area based on the features of different themes and clustering results.

However, some limitations are still in this work which can be improved in the future. Firstly, this work lacks the evolution of this area based on the literature in different periods because of the limited number of documents in this area. Secondly, the current model can not include the analysis of two-word phrases or three-word phrases, which need to be improved in future work by combining the experts' knowledge with the model.

Acknowledgment

The work described in this paper was fully supported by the Research Committee of The Hong Kong Polytechnic University under project code: B-Q57Z and G.45.*.RK2A.

References

Aletras, N., & Stevenson, M. (2013). *Evaluating topic coherence using distributional semantics.* Paper presented at the Proceedings of the 10th International Conference on Computational Semantics (IWCS 2013)–Long Papers.

Ali, M., Mustapha, T. I., Osman, S. B., & Hassan, U. (2020). University social responsibility (USR): An Evolution of the concept and its thematic analysis. *Journal of Cleaner Production*, 124931.

Bastani, K., Namavari, H., & Shaffer, J. (2019). Latent Dirichlet allocation (LDA) for topic modeling of the CFPB consumer complaints. *Expert Systems with Applications, 127*, 256-271.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research, 3*(Jan), 993-1022.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative research in psychology, 3*(2), 77-101.

BROADUS, R. (1987). Toward a definition of "Bibliometrics". Scientometrics. In: Budapest: Akadémiai Kiadó.

Caviggioli, F., & Ughetto, E. (2019). A bibliometric analysis of the research dealing with the impact of additive manufacturing on industry, business and society. *International Journal of production economics, 208*, 254-268.

Chen, Q., Yao, L., & Yang, J. (2016). *Short text classification based on LDA topic model.* Paper presented at the 2016 International Conference on Audio, Language and Image Processing (ICALIP).

Chen, X., Samson, E., Tocqueville, A., & Aubin, J. (2015). Environmental assessment of trout farming in France by life cycle assessment: using bootstrapped principal component analysis to better define system classification. *Journal of Cleaner Production, 87*, 87-95.

Corbett, J., McKeown, P., Peggs, G., & Whatmore, R. (2000). Nanotechnology: international developments and emerging products. *CIRP annals, 49*(2), 523-545.

Darwish, L. R., El-Wakad, M. T., & Farag, M. M. (2021). Towards sustainable industry 4.0: A green real-time IIoT multitask scheduling architecture for distributed 3D printing services. *Journal of Manufacturing Systems, 61*, 196-209.

de Barros Franco, D. G., & Steiner, M. T. A. (2018). Clustering of solar energy facilities using a hybrid fuzzy c-means algorithm initialized by metaheuristics. *Journal of Cleaner Production, 191*, 445-457.

De la Hoz-Correa, A., Muñoz-Leiva, F., & Bakucz, M. (2018). Past themes and future trends in medical tourism research: A co-word analysis. *Tourism Management, 65*, 200-211.

De Oliveira, R. I., Sousa, S. O., & De Campos, F. C. (2019). Lean manufacturing implementation: bibliometric analysis 2007–2018. *The International Journal of Advanced Manufacturing Technology, 101*(1), 979-988.

Ding, Y., Rousseau, R., & Wolfram, D. (2016). *Measuring scholarly impact*: Springer.

Elie, L., Granier, C., & Rigot, S. (2020). The different types of renewable energy finance: A Bibliometric analysis. *Energy Economics*, 104997.

Fan, J., Gao, Y., Zhao, N., Dai, R., Zhang, H., Feng, X., . . . Hambly, B. D. (2020). Bibliometric analysis on COVID-19: a comparison of research between English and Chinese studies. *Frontiers in public health*, *8*, 477.

Gao, S., & Huang, H. (2017). Recent advances in micro-and nano-machining technologies. *Frontiers of Mechanical Engineering*, *12*(1), 18-32.

Gbededo, M. A., Liyanage, K., & Garza-Reyes, J. A. (2018). Towards a Life Cycle Sustainability Analysis: A systematic review of approaches to sustainable manufacturing. *Journal of Cleaner Production*, *184*, 1002-1015.

Gupta, M. K., Song, Q., Liu, Z., Sarikaya, M., Jamil, M., Mia, M., . . . Li, Z. (2020). Ecological, economical and technological perspectives based sustainability assessment in hybrid-cooling assisted machining of Ti-6Al-4 V alloy. *Sustainable Materials and Technologies*, *26*, e00218.

Han, S., Mannan, N., Stein, D. C., Pattipati, K. R., & Bollas, G. M. (2021). Classification and regression models of audio and vibration signals for machine state monitoring in precision machining systems. *Journal of Manufacturing Systems*, *61*, 45-53.

Hartigan, J. A., & Wong, M. A. (1979). Algorithm AS 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C*, *28*(1), 100-108.

Hartini, S., & Ciptomulyono, U. (2015). The relationship between lean and sustainable manufacturing on performance: literature review. *Procedia Manufacturing*, *4*, 38-45.

Helal, M., & Mouhoub, M. (2018). Topic modelling in bangla language: An lda approach to optimize topics and news classification. *Computer and Information Sciences*, *11*(4), 77-83.

Islam, M. T., & Huda, N. (2019). Material flow analysis (MFA) as a strategic tool in E-waste management: Applications, trends and future directions. *244*, 344-361.

Jalali, M. S., Razak, S., Gordon, W., Perakslis, E., & Madnick, S. (2019). Health care and cybersecurity: bibliometric analysis of the literature. *Journal of medical Internet research*, *21*(2), e12644.

Jiang, Q., Liu, Z., Liu, W., Li, T., Cong, W., Zhang, H., & Shi, J. (2018). A principal component analysis based three-dimensional sustainability assessment model to evaluate corporate sustainable performance. *Journal of Cleaner Production*, *187*, 625-637.

Jones, M. V., Coviello, N., & Tang, Y. K. (2011). International entrepreneurship research (1989–2009): a domain ontology and thematic analysis. *Journal of business venturing*, *26*(6), 632-659.

Kanungsukkasem, N., & Leelanupab, T. (2019). Financial latent Dirichlet allocation (FinLDA): Feature extraction in text and data mining for financial time series prediction. *IEEE Access*, *7*, 71645-71664.

Kim, H., Kim, H. K., Kim, M., Park, J., Cho, S., Im, K. B., & Ryu, C. R. (2019). Representation learning for unsupervised heterogeneous multivariate time series segmentation and its application. *Computers and Industrial Engineering*, *130*, 272-281.

Leukel, J., González, J., & Riekert, M. (2021). Adoption of machine learning technology for failure prediction in industrial maintenance: A systematic review. *Journal of Manufacturing Systems, 61*, 87-96.

Li, L., Chen, H., Chen, L., Chen, S., & Shen, Z. (2020). Past and future contributions of artificial reservoirs on global sea-level rise. *Resources, Conservation and Recycling, 161*, 104922.

Liu, G., Gao, X., You, D., & Zhang, N. (2019). Prediction of high power laser welding status based on PCA and SVM classification of multiple sensors. *Journal of Intelligent Manufacturing, 30*(2), 821-832.

Liu, G., Yang, J., Hao, Y., & Zhang, Y. (2018). Big data-informed energy efficiency assessment of China industry sectors based on K-means clustering. *Journal of Cleaner Production, 183*, 304-314.

Luo, L., Tang, Y., Lu, Q., Chen, X., Zhang, P., & Zou, X. (2018). A vision methodology for harvesting robot to detect cutting points on peduncles of double overlapping grape clusters in a vineyard. *Computers in Industry, 99*, 130-139.

Maditati, D. R., Munim, Z. H., Schramm, H.-J., & Kummer, S. (2018). A review of green supply chain management: From bibliometric analysis to a conceptual framework and future research directions. *Resources, Conservation and Recycling, 139*, 150-162.

Malek, J., & Desai, T. N. (2020). A systematic literature review to map literature focus of sustainable manufacturing. *Journal of Cleaner Production, 256*, 120345.

Mani, V., Gunasekaran, A., & Delgado, C. (2018). Supply chain social sustainability: Standard adoption practices in Portuguese manufacturing firms. *International Journal of production economics, 198*, 149-164.

Marino, C. A., & Marufuzzaman, M. (2020). Unsupervised learning for deploying smart charging public infrastructure for electric vehicles in sprawling cities. *Journal of Cleaner Production*, 121926.

Mehrjoo, S., & Bashiri, M. (2013). An application of principal component analysis and logistic regression to facilitate production scheduling decision support system: an automotive industry case. *Journal of Industrial Engineering International, 9*(1), 1-12.

Mohammed, S. H., & Al-augby, S. (2020). LSA & LDA Topic Modeling Classification: Comparison study on E-books. *Indonesian Journal of Electrical Engineering and Computer Science, 19*(1), 353-362.

More, A. (2020). Ultra-Precision Machine Market 2021 : Industry Growth, Competitive Analysis with Top Countries Data, Definition, Market Size, Future Prospects and Forecast to 2026. Retrieved from https://www.marketwatch.com/press-release/ultra-precision-machine-market-2021-industry-growth-competitive-analysis-with-top-countries-data-definition-market-size-future-prospects-and-forecast-to-2026-2020-12-07

Morgan, J., Halton, M., Qiao, Y., & Breslin, J. G. (2021). Industry 4.0 smart reconfigurable manufacturing machines. *Journal of Manufacturing Systems, 59*, 481-506.

Naumanen, M., Uusitalo, T., Huttunen-Saarivirta, E., & van der Have, R. (2019). Development strategies for heavy duty electric battery vehicles: Comparison between China, EU, Japan and USA. *Resources, Conservation and Recycling, 151*, 104413.

Negara, E. S., Triadi, D., & Andryani, R. (2019). *Topic Modelling Twitter Data with Latent Dirichlet Allocation Method.* Paper presented at the 2019 International Conference on Electrical Engineering and Computer Science (ICECOS).

Ocampo-Martinez, C., & Olaru, S. (2020). Dual mode control strategy for the energy efficiency of complex and flexible manufacturing systems. *Journal of Manufacturing Systems, 56*, 104-116.

Pacella, M. (2018). Unsupervised classification of multichannel profile data using PCA: An application to an emission control system. *Computers and Industrial Engineering, 122*, 161-169.

Pang, S., Ban, T., Kadobayashi, Y., & Kasabov, N. K. (2011). LDA merging and splitting with applications to multiagent cooperative learning and system alteration. *IEEE Transactions on Systems, Man, and Cybernetics, Part B, 42*(2), 552-564.

Rajak, S., & Vinodh, S. (2015). Application of fuzzy logic for social sustainability performance evaluation: a case study of an Indian automotive component manufacturing organization. *Journal of Cleaner Production, 108*, 1184-1192.

Roeder, M., Guenther, T., & Zimmermann, A. (2019). Review on fabrication technologies for optical mold inserts. *Micromachines, 10*(4), 233.

Schneider, F., Das, J., Kirsch, B., Linke, B., & Aurich, J. C. (2019). Sustainability in ultra precision and micro machining: a review. *International Journal of Precision Engineering and Manufacturing-Green Technology*, 1-10.

Shi, D., & Tsung, F. (2003). Modelling and diagnosis of feedback-controlled processes using dynamic PCA and neural networks. *International Journal of Production Research, 41*(2), 365-379.

Shiryaev, A. P., Dorofeev, A. V., Fedorov, A. R., Gagarina, L. G., & Zaycev, V. V. (2017). *LDA models for finding trends in technical knowledge domain.* Paper presented at the 2017 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus).

Sodhi, M. S., & Tang, C. S. (2018). Corporate social sustainability in supply chains: a thematic analysis of the literature. *International Journal of Production Research, 56*(1-2), 882-901.

Stoycheva, S., Marchese, D., Paul, C., Padoan, S., Juhmani, A.-s., & Linkov, I. (2018). Multi-criteria decision analysis framework for sustainable manufacturing in automotive industry. *Journal of Cleaner Production, 187*, 257-272.

Sun, L., & Yin, Y. (2017). Discovering themes and trends in transportation research using topic modeling. *Transportation Research Part C: Emerging Technologies, 77*, 49-66.

Syed, S., & Spruit, M. (2017). *Full-text or abstract? examining topic coherence scores using latent dirichlet allocation.* Paper presented at the 2017 IEEE International conference on data science and advanced analytics (DSAA).

Torkayesh, A. E., Malmir, B., & Asadabadi, M. R. (2021). Sustainable waste disposal technology selection: The stratified best-worst multi-criteria decision-making method. *Waste Management, 122*, 100-112.

Van Eck, N. J., & Waltman, L. (2010). Software survey: VOSviewer, a computer program for bibliometric mapping. *Scientometrics, 84*(2), 523-538.

Van Eck, N. J., & Waltman, L. (2013). VOSviewer manual. *Leiden: Univeristeit Leiden, 1*(1), 1-53.

Van Eck, N. J., Waltman, L., Dekker, R., & van den Berg, J. (2010). A comparison of two techniques for bibliometric mapping: Multidimensional scaling and VOS. *Journal of the American Society for Information Science Technology, 61*(12), 2405-2416.

Vicente, P., & Reis, E. (2007). Segmenting households according to recycling attitudes in a Portuguese urban area. *Resources, Conservation and Recycling, 52*(1), 1-12.

Wang, G., Zhang, Y., Liu, C., Xie, Q., & Xu, Y. (2019). A new tool wear monitoring method based on multi-scale PCA. *Journal of Intelligent Manufacturing, 30*(1), 113-122.

Wang, Z., Zhang, W., & Zhou, H. (2019). Perception-guided multi-channel visual feature fusion for image retargeting. *Signal Processing: Image Communication, 79*, 63-70.

Xiong, H., Cheng, Y., Zhao, W., & Liu, J. (2019). Analyzing scientific research topics in manufacturing field using a topic model. *Computers and Industrial Engineering, 135*, 333-347.

Yip, W., & To, S. (2020). Sustainable ultra-precision machining of titanium alloy using intermittent cutting. *Journal of Precision Engineering and Manufacturing-Green Technology, 7*(2), 361-373.

Yip, W., To, S., & Zhou, H. (2020). Social network analysis for optimal machining conditions in ultra-precision manufacturing. *Journal of Manufacturing Systems, 56*, 93-103.

Yip, W. S., To, S., & Zhou, H. (2021). Current status, challenges and opportunities of sustainable ultra-precision manufacturing. *Journal of Intelligent Manufacturing*, 1-13.

Zhou, H., Yip, W. S., Ren, J., & To, S. (2020). An Interaction Investigation of the Contributing Factors of the Bullwhip Effect Using a Bi-Level Social Network Analysis Approach. *IEEE Access, 8*, 208737-208752.

Zhou, R., Awasthi, A., & Stal-Le Cardinal, J. (2020). The main trends for multi-tier supply chain in Industry 4.0 based on Natural Language Processing. *Computers in Industry*, 103369.