Journal of Water Resource and Protection, 2024, 16, 450-465
https://www.scirp.org/journal/jwarp
ISSN Online: 1945-3108
ISSN Print: 1945-3094

Study Progress Analysis of Effluent Quality Prediction in Activated Sludge Process Based on CiteSpace

Kemeng Xue

Beijing No.4 High School International Campus, Beijing, China
Email: mollyxkm@126.com

How to cite this paper: Xue, K.M. (2024) Study Progress Analysis of Effluent Quality Prediction in Activated Sludge Process Based on CiteSpace. *Journal of Water Resource and Protection*, **16**, 450-465.
https://doi.org/10.4236/jwarp.2024.166025

Received: April 22, 2024
Accepted: June 18, 2024
Published: June 21, 2024

Abstract

In this paper, CiteSpace, a bibliometrics software, was adopted to collect research papers published on the Web of Science, which are relevant to biological model and effluent quality prediction in activated sludge process in the wastewater treatment. By the way of trend map, keyword knowledge map, and co-cited knowledge map, specific visualization analysis and identification of the authors, institutions and regions were concluded. Furthermore, the topics and hotspots of water quality prediction in activated sludge process through the literature-co-citation-based cluster analysis and literature citation burst analysis were also determined, which not only reflected the historical evolution progress to a certain extent, but also provided the direction and insight of the knowledge structure of water quality prediction and activated sludge process for future research.

Keywords

Biological Model, Effluent Quality Prediction, Activated Sludge Process, CiteSpace, Knowledge Map, Co-Citation Cluster Analysis

1. Introduction

The scientific knowledge graph is a bibliometric method first proposed by the National Academy of Sciences in 2003, which focuses on scientific knowledge as a quantitative research object. Through the drawing of the scientific knowledge graph, citation analysis is combined with data and information visualization, and complex scientific knowledge fields are visualized through data mining, information processing, knowledge measurement, and graphic drawing. The scientific knowledge graph can visualize the development process and structural relationships of scientific knowledge, reveal the laws of scientific knowledge and its activities and demonstrate the scientific relationship and evolution laws of knowledge structure. The specific methods in scientific knowledge graph include citation analysis, co-citation analysis, co-word analysis, clustering analysis, word frequency analysis, social network analysis, multidimensional scale analysis, and so on.

CiteSpace is currently an internationally recognized software for drawing scientific knowledge graphs [1] [2] [3] [4] [5]. It is a scientific literature visualization and analysis software developed in Java language, which can display information such as authors, keywords, institutions in a certain field using scientific knowledge graphs. It is an important tool for drawing scientific knowledge graphs at present.

Water quality prediction is an important application of automatic control technology in the field of environmental protection, especially in the treatment of wastewater. Wastewater treatment is a complex nonlinear process involving physical, chemical, and biological processes. Although automatic water quality detection instruments have been applied to a certain extent, they can only summarize the situation that has already occurred and cannot predict future trends or provide early warning for possible events. The existing ASM models in the field of sewage treatment have complex modeling processes, and the assumptions and actual deviations are far from reality. The difficulty in solving multi-component nonlinear equation systems is an important factor limiting the application of this series of ASM models. At the same time, sewage treatment technologies are diverse and the principles are not the same. Therefore, the complex physical, chemical, and biological processes, as well as the practical application of existing models, are important challenges facing water quality prediction at present.

The activated sludge process is a typical method of wastewater treatment, discovered by Clark and Gage in the UK in 1912. The basic principle of activated sludge process is the energy demand of suspended microbial communities during growth and metabolism processes in wastewater, which requires mathematical models to understand deeply. On the other hand, computers, big data technology, mathematical simulation and artificial intelligence algorithms have more and more application in the fields of water management and environmental protection. Big data technology is an important method for abstracting regular data from industrial processes, especially suitable for industrial processes with relatively stable processes, especially the activated sludge process in popular sewage treatment. In big data technology, statistical analysis methods, neural network-based methods, fuzzy mathematics methods, etc., have been applied. In statistical analysis methods, based on the normalization analysis of continuous online monitoring of water quality indicators, for sewage plants that are basically stable in operation, it is possible to predict relatively close trend data and operating boundary conditions, and thus provide early warning for potential water quality unqualified accidents.

At the same time, the operation of sewage treatment plants requires a large amount of data support through network and automation equipment (Wang Shuaishuai, 2024). Although online monitoring equipment can present the current situation and historical data, it cannot forecast the future change trend of effluent. Therefore, water quality prediction becomes a very critical issue to calculate and forecast the future trend of water quality using actual historical data and mathematical models. It can forecast the future trend of water quality and take some measures in advance to prevent water pollution.

In summary, the study aims to Modeling and analyzing the growth and metabolism of microorganisms in the activated sludge process, as well as predicting the effluent quality, play important roles to improve the treatment process and ensure the stable operation of sewage plants. The international research progress should be paid attention and learned deeply.

In this paper, CiteSpace software was adopted to quantitatively analyze the knowledge graph of international research papers on water quality prediction and activated sludge process. We can then accurately grasp the key ideas of foreign literature on the same time dimension, and point out new research hotspots and future trend in the field of water quality prediction. It can conclude the depth and breadth of research and provide effective insights for the future development.

2. Methods

2.1. CiteSpace Introduction and Version

The scientific knowledge graph displays the development process and structural relationships of complex scientific knowledge through data mining, information processing, knowledge measurement, and graphic drawing, revealing scientific knowledge and its activity patterns, and displaying the structural relationships and evolution laws of knowledge.

The focus of the scientific knowledge graph is on the application of visualization techniques that combine citation visualization, information visualization, and computer technology in literature analysis. The research methods mainly include citation analysis, co-citation analysis, co-word analysis, cluster analysis, word frequency analysis, social network analysis, and multidimensional scale analysis.

The version of CiteSpace in this paper is 6.2.R2. CiteSpace's analysis mainly involves the following aspects [6] [7] [8].

(1) Co-occurrence analysis

In a certain case, nodes and cooperative links between nodes can form a co-occurrence graph. Keyword co-occurrence analysis is a process based on co-occurrence graph. The number and size of nodes and the number of cooperative links between nodes can be drawn as intuitive evidence for co-occurrence analysis.

Generally, a circular node represents a keyword. The different color of the nodes' annual ring structure represents the time when a keyword appears. The size of a node depends on the occurrence frequency and the larger the node is, the higher the frequency of keyword occurrence is. The thicker the cooperative links between nodes are, the closer the connection between keywords are, and the higher the frequency of keywords occurrence are.

(2) Cluster analysis

Keyword clustering is a clustering process which closely related keywords into one category, and it can clearly indicate the research hotspots in a certain academic field or technology. Generally, module value Q and average contour value S are the main indicators in cluster analysis.

(3) Timeline analysis

The main function of timeline analysis is to analyze the trend of keywords in each clustering module over time. It can reveal the evolution law of a certain technology and the relationship between clusters in a certain field. The more nodes in a cluster are, the more important the clustering fields are.

2.2. Terms Definition and Data Extraction

We choose the publications of the Web of Science-Science Citation Index Expanded (SCIE) and Social Science Citation Index (SSCI) (WoS) as the data source. WoS is the premier, global, highly recognized, and reliable academic publication database with multiple scientific fields. Specifically, we choose SCI and SSCI because they are the core collections of WoS; publications in SCI and SSCI enjoy higher quality and reputation compared with other databases or collections, such as Engineering Index (EI) and Emerging Science Citation Index (ESCI). Furthermore, they cover all the research topics among global scholars, making research results more convincing and universal. Therefore, the results of this study are authoritative, appropriate, and representative.

We collect the data until Feb 15, 2024, using the following search formula.

(1) Select Web of Science Core Collection—SCIE and SSCI.

(2) Topic = ("quality prediction") AND Topic = ("activated sludge process").

(3) Language = "English", and the period = "2007 - 2024" (index date from 2007-01-01 to 2024-02-15).

(4) Document type = "Article" or "Review".

The SCIE and SSCI we adopt cover representative research globally on quality prediction and activated sludge process. Besides, we restrict the document language to English for unified and practical analysis. Finally, 103 publications (data) for analysis.

There are several steps in this study as below:

(1) Statistical analysis: including the temporal distributions of publications on quality prediction and activated sludge process and the category classification of the publications. The statistical characteristics can display the knowledge foundation of quality prediction and activated sludge process research.

(2) Collaboration analysis: including the author, institution, and region collaboration network analysis. Collaboration networks reveal the correlations of

quality prediction and activated sludge process from multiple perspectives.

(3) Co-occurrence analysis: keywords co-occurred networks can reveal potential hotspots and trends for quality prediction and activated sludge process research.

(4) Timeline analysis: analyzing the trend of keywords in each clustering module over time.

3. Results and Discussions

3.1. Statistical Analysis

The annual distribution of publications aims to help us understand the research outputs on quality prediction and activated sludge process. It explains the dynamics of quality prediction and activated sludge process research in the past and helps scholars judge the future development status. Number of publications and citations by years from 2007 to 2024 was shown in Figure 1. We can find that the number of journal publications and citations were increasing with fluctuations.

From the bar chart of publications number, it can be seen that the number of publications steadily increased from 2007 to 2012, reaching the first peak in 2012. Between 2012 and 2017, there was a fluctuation, and it was not until 2017 that the second peak was reached in the number of publications. Since 2018, the number of publications has remained relatively stable, reaching the third peak only in 2022. The number of publications in other years has remained relatively stable.

From the line chart of citation number, it can be seen that the trend of citation number is basically similar to the publication number. Before 2022, it was in a steady upward trend, and after 2022, citation number decreased significantly. In conclusion, the rising trend proves that research about quality prediction and activated sludge process is still prevalent.

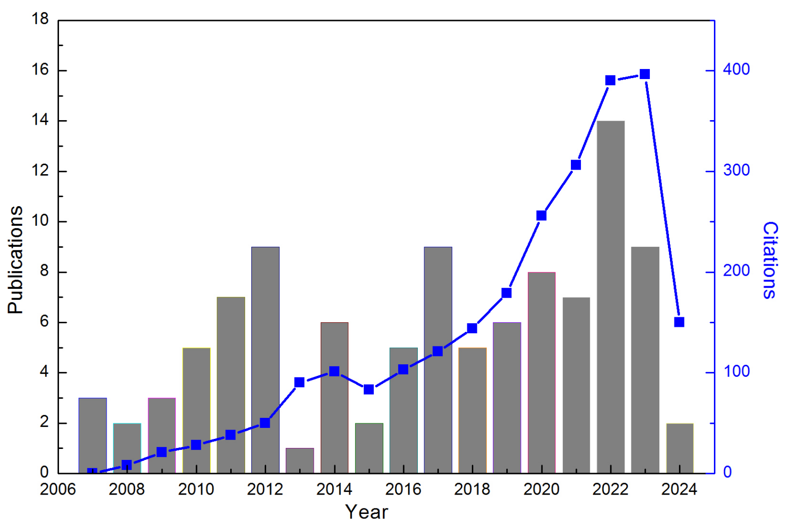

Figure 1. Number of publications and citations by years from 2007 to 2024.

3.2. Collaboration Analysis

The collaboration networks are to understand the links between different authors, institutions, and regions on quality prediction and activated sludge process. Analyzing these collaboration links can help scholars understand the current research correlations and seek potential cooperation partners. This study will especially investigate the core strength of the authors, institutions, and regions in the effluent quality prediction and activated sludge process. Based on completing the analysis of absolute values of publication number and citation number in Section 2.1, this paper focused on three factors: authors, institutions, and regions on quality prediction and activated sludge process.

3.2.1. Author Collaboration Network

The analysis process for the author was as follows: firstly, in CiteSpace software, the search time span was set to 2007 to 2024, with a single time interval of 1 year. All literature searched by the author was arranged, and authors who have published more than or equal to 2 papers were listed in Table 1 in descending order. Then a network relationship diagram between them was drawn as shown in Figure 2.

Table 1. Author collaboration for more than one paper.

Rank	Number of Papers	Centrality	Year	Authors
1	6	0.01	2009	Flores-alsina, Xavier
2	6	0.01	2008	Vanrolleghem, Peter A
3	4	0.00	2018	Barbusinski, Krzysztof
4	3	0.00	2020	Du, Xianjun
5	3	0.00	2009	Gernaey, Krist V
6	3	0.00	2022	Gernaey, Krist
7	3	0.00	2010	Benedetti, Lorenzo
8	3	0.00	2019	Dai, Hongliang
9	2	0.00	2022	Monje, Vicente
10	2	0.00	2018	Liu, Zheng
11	2	0.00	2014	Jeppsson, Ulf
12	2	0.00	2019	Chen, Wenliang
13	2	0.00	2022	Junicke, Helena
14	2	0.00	2022	Kjellberg, Kasper
15	2	0.00	2017	Al-omari, Ahmed
16	2	0.00	2010	Mannina, Giorgio
17	2	0.00	2018	Han, Hong-Gui
18	2	0.00	2019	He, Lei
19	2	0.00	2020	Szelag, Bartosz
20	2	0.00	2009	Rodriguez-roda, Ignasi
21	2	0.00	2010	Nopens, Ingmar
22	2	0.00	2016	Batstone, Damien J

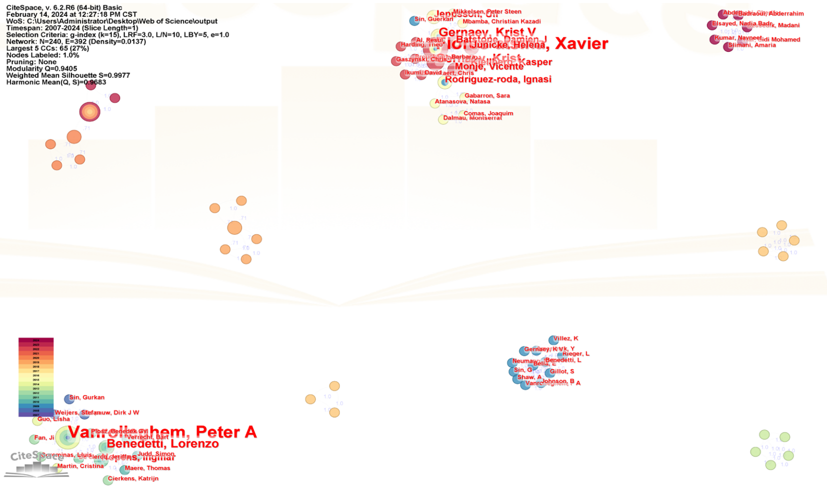

Figure 2. Author collaboration network.

The results of **Table 1** indicated that Flores alsina Xavier, Vanrolleghem Peter A, and Barbusinski Krzysztof have published more papers, ranking in the top three in **Table 1** with 6, 6, and 4 papers respectively. Only 22 authors have published more than one paper, indicating that there are not many scholars who have repeatedly studied "water quality prediction" and "activated sludge process".

The collaboration network map of the author in **Figure 2** showed a total of 240 nodes, 392 cooperative links, and a network density of 0.0137. The module value (Q value) was 0.9405 (>0.3), indicating that the research team structure was more significant because the Q value was larger than 0.3. The average contour value (S value) was 0.9977 (>0.7). Generally, if $S > 0.5$, it indicated that clustering was reasonable. In this conclusion, because $S > 0.7$, it indicated that the clustering analysis of the significant graph was very reliable. In the above analysis, the average contour values of the graph were larger than 0.7, indicating that the clustering cluster studied had high credibility. The module values of the data clustering indicated that the divided author team structure was significant, the network module performance was good, and the clustering analysis was reasonable.

According to the overview in Section 1.1, the larger the node were, the more papers the author had published. The author group consisted of Flores alsina Xavier, Vanrolleghem Peter A, Barbusinski Krzysztof, and others as core nodes, forming a relatively clear network of interrelationships, with rich research context and structure inside. However, the intersection between clusters was not similar, indicating that most research teams or individuals were more focused on individual research results, or only cooperate with the group of scholars within the team. The cooperation between teams was insufficient, and in the long run, cooperation between different teams should be more advocated.

3.2.2. Institution Collaboration Network

Similar to the author's retrieval method, institutions with more than one paper published were shown in **Table 2**. Among them, Technical University of Denmark has the highest number of publications, with 9 papers published. Secondly, institutions such as Laval University, Kierce University of Technology, Harbin Institute of Technology, and Ghent University also have a significant number of publications, with 8, 6, 6, and 6 papers respectively. Institutions with a large number of publications have conducted more research on “water quality prediction” and “activated sludge method”, and their research depth and breadth are relatively complete. Furthur more, we can find that most of the top 31 collaborative institutions are located in Europe (18 institutions), and Asia (9 institutions), indicating quality prediction and activated sludge process was a highly collaborated topic in these continents. National environmental and energy policies influence institutions’ cooperation. For example, the national authorities of Denmark and Finland pay special attention to environmental protection; the national authority in China initiated abundant regulations to support new energy source development, leading to more institutional cooperation in this field.

Table 2. Institution collaboration for more than one paper.

Rank	Number of Papers	Centrality	Year	Authors	Continents
1	9	0.03	2008	Technical University of Denmark	Europe
2	8	0.15	2008	Laval University	America
3	6	0.00	2016	Kielce University of Technology	Europe
4	6	0.00	2017	Harbin Institute of Technology	Asia
5	6	0.04	2008	Ghent University	Europe
6	5	0.00	2012	Beijing University of Technology	Asia
7	5	0.02	2012	Lund University	Europe
8	3	0.09	2019	Jiangsu University of Science and Technology	Asia
9	3	0.00	2020	Lanzhou University of Technology	Asia
10	3	0.00	2009	Universitat de Girona	Europe
11	3	0.06	2012	Institut Catala de Recerca de I’Aigua (ICRA)	Europe
12	3	0.00	2011	Kyung Hee University	Asia
13	2	0.00	2009	Universite Paris Saclay	Europe
14	2	0.00	2011	South China University of Technology	Asia
15	2	0.00	2020	Lublin University of Technology	Europe
16	2	0.00	2012	Helmholtz Center for Evironmental Research (UFZ)	Europe
17	2	0.00	2016	University of Queenland	Australia
18	2	0.00	2018	Systems Research Institute of the polish Academy of Science	Europe
19	2	0.06	2021	Huazhong University of Science and Technology	Asia
20	2	0.09	2012	East China University of Science and Technology	Asia
21	2	0.00	2018	Polish Academy of Sciences	Europe
22	2	0.01	2009	CH2M Hill	America
23	2	0.00	2009	INRAE	Europe
24	2	0.00	2012	Helmholtz Association	Europe
25	2	0.00	2009	Swiss Federal Institute of Aquatic Science and Technology	Europe
26	2	0.00	2018	Silesian University of Technology	Europe
27	2	0.00	2010	University of Palermo	Europe

Continued

28	2	0.03	2021	Egyptian Knowledge Bank (EKB)	Africa
29	2	0.00	2022	Nonozymes	Europe
30	2	0.00	2012	Pusan National University	Asia
31	2	0.01	2010	Cranfield University	Europe

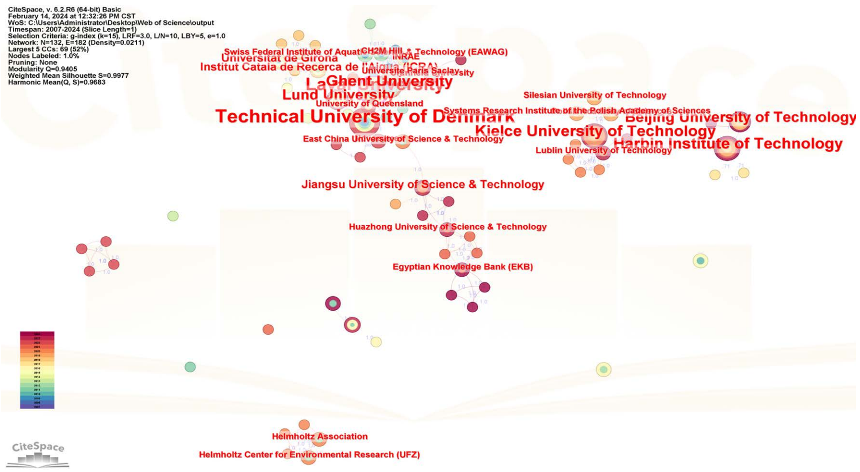

Figure 3. Institution collaboration network.

The knowledge graph analyzed based on the number of institutional publications was shown in Figure 3. The graph showed a total of 132 nodes, 182 cooperative links, a network density of 0.0211, a module value (Q value) of 0.9405, and an average contour value (S value) of 0.9977. The analysis results of Q and S values also indicate that the clustering analysis has high credibility and is reasonable.

The distribution of institutional communities with Technical University of Denmark, Kierce University of Technology, Ghent University, and Lund University as the core was relatively concentrated, and the internal cooperative links of institutional groups were close, indicating that research institutions had formed a clear network of relationships when studying "water quality prediction" and "activated sludge method", promoting further communication and exchange of the same research topic.

However, the centrality of these top collaborated institutions is generally low; the centrality reveals the collaboration strength or frequency among the nodes, and higher centrality implies more vital collaboration or associations between the observed nodes and other nodes. This finding implies that institutional collaboration research still needs to be higher, and more institutional cooperation should be encouraged. The low centrality is because national laws, regulations, and the social and natural environment affect research and application; simultaneously, different regions have specialized characteristics, leading to fewer research interest overlaps. That again confirms that different regions have different main focuses on, leading to different cooperation focus of institutions. Thus, scholars from different institutions should look for common research interests of other institutions and initiate potential cooperation.

3.2.3. Region Collaboration Network

Like the author and institution search methods, the regions where the total number of published papers with more than 1 were shown in Table 3. The region with the highest number of publications was China, with a total of 38 articles. Secondly, regions such as USA, Canada, Denmark, and South Korea had a relatively large number of articles, with 12, 10, 10, and 7 articles respectively. There were numerous research papers in regions with a large number of publications, and the depth and breadth of research were relatively complete, making a significant contribution to the core of this field.

The knowledge graph analyzed based on national publication volume was shown in Figure 4. Among them, the graph showed a total of 40 nodes, 79 connecting lines, a network density of 0.1013, a module value (Q value) of 0.8141, and an average contour value (S value) of 0.9797. The analysis results of Q and S values also indicated that the clustering analysis had high credibility and was reasonable.

Table 3. Region collaboration for more than one paper.

Rank	Number of Papers	Centrality	Year	Authors
1	38	0.41	2011	China
2	12	0.24	2017	USA
3	10	0.21	2008	Canada
4	10	0.25	2008	Denmark
5	7	0.01	2011	South Korea
6	6	0.21	2008	Belgium
7	6	0.09	2007	France
8	6	0.08	2009	Spain
9	6	0.02	2010	Italy
10	6	0.08	2016	Poland
11	5	0.04	2012	Swenden
12	5	0.01	2013	Australia
13	5	0.00	2011	Iran
14	4	0.08	2010	Germany
15	4	0.06	2010	England
16	4	0.07	2011	Scotland
17	3	0.00	2009	Greece
18	3	0.00	2019	India
19	2	0.00	2009	Turkey
20	2	0.00	2015	Slovenia
21	2	0.14	2021	Egypt
22	2	0.00	2009	Switzerland
23	2	0.03	2011	Algeria
24	2	0.00	2012	Netherlands

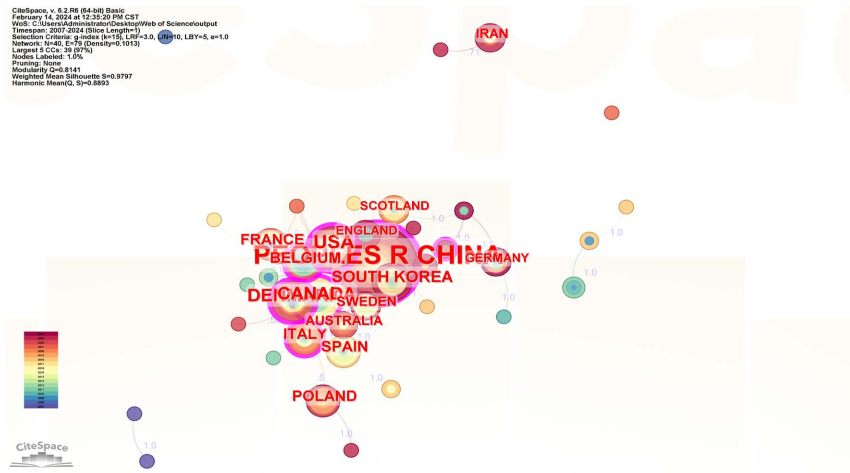

Figure 4. Region collaboration network.

The distribution of region cluster centered around China was very concentrated, and the internal cooperative links of institutional groups are very close, indicating that the regions studied in "water quality prediction" and "activated sludge method" have formed a relatively clear network of relationships, promoting further communication and exchange of the same research topic. This field has a more unified research direction in centralized learning, a closer thematic exploration area, and deep cooperative links between regions, which is conducive to more comprehensive and complete academic exploration in the future.

3.2.4. Keywords Co-Occurrence

Table 4 showed the core keywords whose number of papers was greater than or equal to 4. Several keywords with large number of publications were activated sludge, removal, performance, model, activated sludge process, waste water, prediction, which were 28, 21, 15, 15, 15, 15, 14, and 14 papers respectively. To some extent, it showed that the core research of knowledge map was carried out around these main keywords, reflecting that the core vocabulary of water quality prediction and activated sludge process had good inheritance.

Keyword expressed the research theme of the paper, which was the essence and core of the research. CiteSpace was used to draw the network relationship map of high-frequency keywords, and the knowledge graph was analyzed by the closeness between keywords. CiteSpace analyzed the topics of water quality prediction and activated sludge process with a time span of 2007 - 2024. A single time zone was divided into 1 year. The 25 keywords with the highest citation frequency in each time zone of 1 year were extracted, and each keyword generates over 4 papers. Among them, the knowledge map had 248 keyword nodes, 1074 cooperative links. The network density was 0.0351.

Figure 5 was the knowledge map of keywords. Seven keyword populations were formed according to the classification of different keywords. The first eight keywords were neural network, asm2d-guided reward, decision tree, sludge process, wastewater treatment modelling, smart systems functioning, transformation product, uncertainty analysis, full-scale industrial waste treatment plant. Many nodes were formed in the interleaved map, and the difference of node sizes also represented the different degrees of keyword occurrence frequency. The larger the nodes were, the more times the keywords appeared. Those 248 keywords in the whole figure were connected by nodes of different colors in a complex interleaving pattern. In addition, the thicker the lines, the greater the correlation between keywords. The smaller the clustering values were, the more high-frequency keywords were studied, and the closer the surrounding lines were. On the contrary, the fewer the number of keyword studied, the more they were distributed in the edge region of the graph, and the surrounding lines were not dense, and there were not so many nodes. The internal connection of the key words group was very close, indicating that the related research topics of water quality prediction and activated sludge process were rich and interrelated, which reflect the large number and rich types of related research were extremely important.

Table 4. Keywords results for more than four papers.

Rank	Number of Papers	Centrality	Year	Keywords
1	38	0.75	2007	activated sludge
2	21	0.20	2010	removal
3	15	0.20	2007	performance
4	15	0.16	2008	model
5	15	0.14	2008	activated sludge process
6	14	0.12	2010	waste water
7	14	0.16	2008	prediction
8	8	0.11	2010	waste water treatment
9	8	0.03	2019	machine learning
10	7	0.07	2014	simulation
11	6	0.09	2009	design
12	6	0.02	2019	system
13	6	0.09	2009	calibration
14	6	0.07	2007	activated sludge models
15	5	0.10	2013	degradation
16	5	0.02	2020	wastewater treatment plant
17	5	0.03	2012	control strategy
18	5	0.03	2008	artificial neural network
19	4	0.02	2019	sludge
20	4	0.01	2007	membrane bioreactor
21	4	0.00	2011	artificial neural networks
22	4	0.04	2013	activated sludge model
23	4	0.00	2008	optimization
24	4	0.02	2022	recovery
25	4	0.02	2018	treatment plant

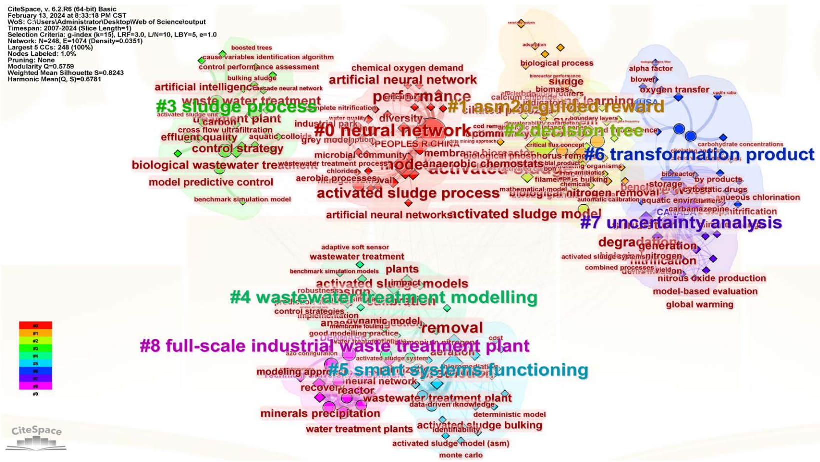

Figure 5. The keywords co-occurrence network.

3.2.5. Timeline Graph Analysis

(1) Timeline graph of authors, institutions and regions

Timeline of Author, institution, and region were classified in Figure 6 according to keywords from 2007 to 2024. Figure 6 showed the temporal changes of author, institution, and region with 742 nodes and 3497 links.

The timeline analysis of the author showed that Gernaey Krist began researching model simulation before 2010 and continued to study it until 2020. Scholars have been studying this keyword for the past decade, and the cooperative links between nodes indicated that the research of multiple scholars were related, leading to an increasing size of the first node. Reuse was studied in 2007 as Flores alsina Xavier, and the number of studies gradually decreased until 2015, during which many authors were interrelated to study reuse. After 2015, the number of scholars studying the reuse keyword gradually decreased, and the number of nodes on the timeline gradually decreased.

The timeline analysis of institutions showed that Laval University had been conducting research on modeling since 2007, Ghent University had been conducting research on modeling, and Technical University of Denmark had been conducting research on bacteria. Research on various keywords has been ongoing until 2024. The remaining keywords began to be studied gradually after 2010, and all keyword themes have been studied for a long time in various institutions, reflecting the depth of attention paid to "water quality prediction" and "activated sludge method" by the aforementioned schools.

Timeline analysis of regions shows that since 2007, Canada, Denmark, Spain have been conducting research on Activated sludge, France, USA, Belgium and other regions have been conducting research on BAF, and China has started a significant amount of publications and research on model predictive. The number of national research articles on other keywords is relatively small, and the frequency of national research has gradually decreased over time, indicating that the research focus of each region on "water quality prediction" and "activated sludge method" is basically based on the top four keywords before 2015.

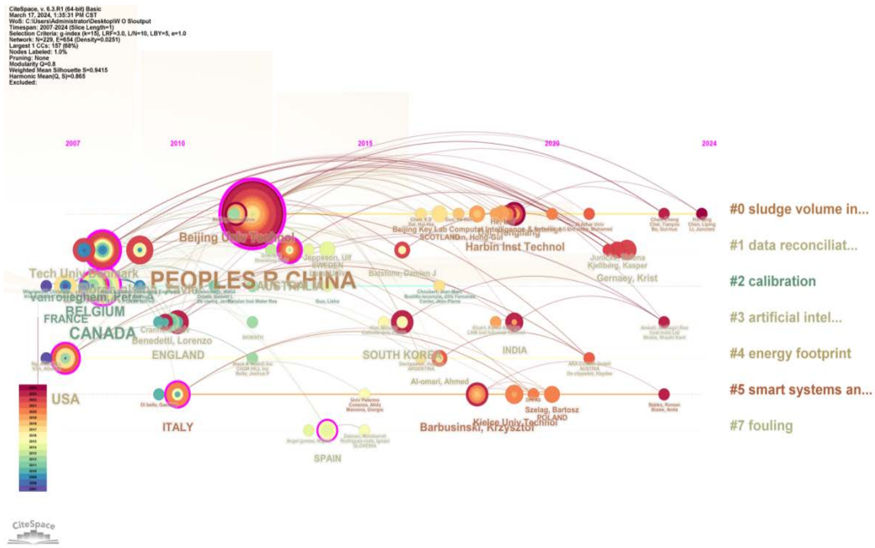

Figure 6. Timeline of author, instution, and region co-occurrence.

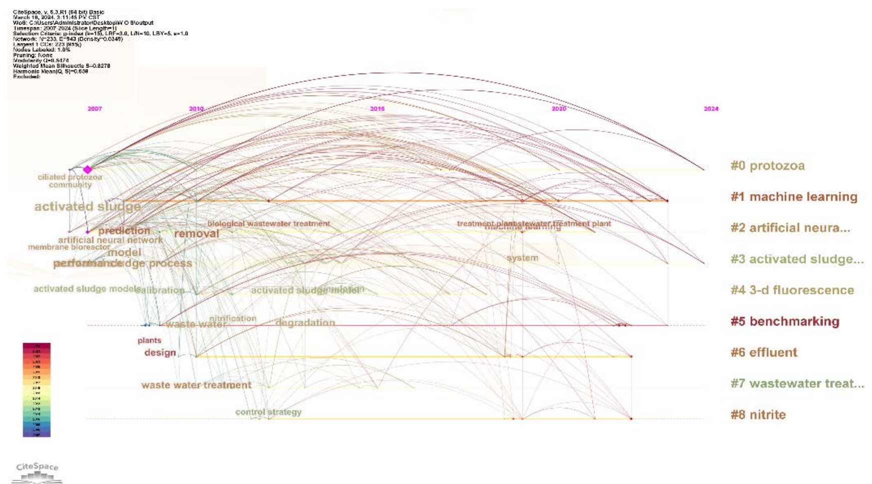

Figure 7. Timeline graph analysis of keyword knowledge.

(2) Timeline graph of key words

Keywords have been extracted and classified according to time period from 2007 to 2024, as shown in Figure 7. Since 2007, activated sludge, ciliated protozoa community, waste water treatment and other keywords have been studied, and each keyword will be continuously studied until 2020 and beyond.

The most recent cluster is #1, "machine learning" in quality prediction and activated sludge process research, demonstrating its continuous popularity among scholars. These representative scholars evaluate data on artificial intelligence in quality prediction and activated sludge process to improve automation. Besides, it is worth noting that "machine learning", is one of the latest clusters, demonstrating that it is an emerging research topic in recent years. Machine learning has been used to evaluate to predict operational status and ensure their stability and reliability.

Although the number of published papers has decreased with the evolution of time, each study is related to the former, forming a dense curve connection model, which can make the research in this field persist into the future and promote the further development of water quality prediction and activated sludge method. There are also close cooperative links between keywords, indicating that there are interrelated research subjects in topics and topics, and the cooperation between individuals and teams provides stable solutions and practice plans for the future strategic development of environmental governance.

4. Conclusions

In this paper, CiteSpace was adopted to analyze the data on "water quality prediction" and "activated sludge peocess" research in the core journal library included in the WOS database. The author, institution, region collaboration network and Keywords Co-occurrence of the 17-year research are drawn. Through graphic analysis, the following conclusions can be drawn:

(1) From 2007 to 2024, a total of 103 articles were retrieved. As for the author, there were only 22 authors who published more than one article, indicating that there were not many scholars who repeatedly studied "water quality prediction" and "activated sludge method". As for institution, the ones with a higher volume of publications are Technical University of Denmark, Laval University, Kielce University of Technology, Harbin Institute of Technology, and Ghent University. As for region, China, USA, Canada, and Denmark are the ones with more published papers.

(2) Timeline Graph Analysis based on author, institution, and region based on keywords from 2007 to 2024, showed that all institutions and authors continued to conduct research work. Since 2007, keywords such as activated sludge, ciliated protozoa community, and waste water treatment have been under research, and each keyword has been continuously studied until 2020. Although the number of publications has decreased recently, each study was sustainable which persists into the future and promotes further development of "water quality prediction" and "activated sludge method".

Conflicts of Interest

The author declares no conflicts of interest regarding the publication of this paper.

References

[1] Li, H., Xiang, Y., Yang, W., Lin, T., Xiao, Q. and Zhang, G. (2024) Green Roof Development Knowledge Map: A Review of Visual Analysis Using CiteSpace and VOSviewer. *Heliyon*, **10**, e24958. https://doi.org/10.1016/j.heliyon.2024.e24958

[2] Li, Y.-Q., Yao, J.-P., Yan, X.-Y., Wang, W., Zhang, Y.-P., Yuan, L., *et al.* (2024) Advances in Acupuncture Regulation on the Autonomic Nervous System from 2013 to

2022: A Bibliometric Analysis via CiteSpace. *Complementary Therapies in Medicine*, **80**, Article 103009. https://doi.org/10.1016/j.ctim.2023.103009

[3] Geng, Y., Zhang, N. and Zhu, R. (2023) Research Progress Analysis of Sustainable Smart Grid Based on CiteSpace. *Energy Strategy Reviews*, **48**, Article 101111. https://doi.org/10.1016/j.esr.2023.101111

[4] Bamisile, O., Zheng, Z., Adun, H., Cai, D., Ting, N. and Huang, Q. (2023) Development and Prospect of Flywheel Energy Storage Technology: A CiteSpace-Based Visual Analysis. *Energy Reports*, **9**, 494-505. https://doi.org/10.1016/j.egyr.2023.05.147

[5] Xiao, F., Li, C., Sun, J. and Zhang, L. (2017) Knowledge Domain and Emerging Trends in Organic Photovoltaic Technology: A Scientometric Review Based on CiteSpace Analysis. *Frontiers in Chemistry*, **5**, 67. https://doi.org/10.3389/fchem.2017.00067

[6] Zhao, X., Nan, D., Chen, C., Zhang, S., Che, S. and Kim, J.H. (2023) Bibliometric Study on Environmental, Social, and Governance Research Using CiteSpace. *Frontiers in Environmental Science*, **10**, Article 1087493. https://doi.org/10.3389/fenvs.2022.1087493

[7] Wang, S.S., Liu, J., Han, W.S., *et al.* (2024) Analysis on Research Hot Spots of New Energy Vehicle Technology Based on CiteSpace. *Automobile Applied Technology*, 8.

[8] Tian, S.W., Zhang, X.Y., Yang, Q., *et al.* (2024) Comparative and Visualization Analysis of Research Hotspots and Development Trends in Tooth Loss Studies at Home and Abroad Based on CiteSpace. *Journal of Clinical Stomatology*, 112.