Analyzing New Power System Scholarly Research Trends Based on CNKI Via CiteSpace

Yanbo Wang[1,a], Yingkun Han[2,b], Yan Ma[2,c], Zhuo Yang[3,d*], Tongxuan Chen[3,e]

[1]Global Energy Interconnection Group Co. Ltd, Beijing, China
[2]State Grid Shandong Electric Power Research Institute, Jinan, China
[3]North China Electric Power University, Beijing, China

[a]yanbo-wang@geidco.org, [b]hanyk8086@163.com, [c]yanpony@126.com
[d*]Corresponding author's e-mail: 408137867@qq.com
[e]916431042@qq.com

Abstract. The new power system is the key to energy and power transformation, and is of great significance to the realization of sustainable energy development. Scientific papers can reflect the development trend of technology and theory in a certain field and guide the future research direction. The purpose of this paper is to utilize text co-occurrence, cluster analysis and bibliometric analysis to explore the research hotspots and development trends of new power system. Firstly, a literature dataset of "new power system" was established based on CNKI database. Then, CiteSpace visual analysis software was used to analyze the keyword co-occurrence and cluster analysis of the literature dataset, and summarize the keyword clusters that can describe the new electric power system. Then, the keyword clusters were used as search fields to search the CNKI database and analyze the trend of publications in the fields described by the keyword clusters, so as to reveal the overall development trend of the research on new power systems. This study helps to clarify the current status of research on new power systems, grasp its development trend, and provide guidance for academic research and technological innovation in the field of energy and power. At the same time, the proposed research method can also be applied to the bibliometric analysis of other fields.

Keywords: New Power System, Cluster Analysis, Data Visualization, Bibliometrics

1 Introduction

The new type of power system is to guarantee energy and power security [1] as the basic premise, to meet the economic and social development of high-quality power demand as the primary goal, to build a high proportion of new energy supply and demand system as the main task [2], to the source, network, load, storage [3] multi-directional synergies, flexible interactions [4] as a strong support for a strong, intelligent and flexible grid as a pivotal platform, scientific and technological innovation and institutional

Y. Kuang et al. (eds.), *Proceedings of the 2024 5th International Conference on Education, Knowledge and Information Management (ICEKIM 2024)*, Atlantis Highlights in Computer Sciences 22,
https://doi.org/10.2991/978-94-6463-502-7_63

innovation and institutional mechanism innovation as the fundamental guarantee of the new era of power system. Fundamental guarantee of the new era of power system. It is an important part of the new energy system and a key carrier for solving the global energy crisis [5]. As a theoretical carrier of science and technology [6], scientific and technical papers often reflect the development trend of technical theories in a certain field, guide the research direction of researchers, and ultimately bring about an innovative technological revolution. Therefore, it is necessary to analyze scientific and technical papers through methods such as bibliometrics or big data analysis [7] to tap the research trends in a certain field.

In this paper, focusing on the current literature data related to new power system technology, we establish the 'new power system' dataset required for the research based on the CNKI database, and conduct data visualization and analysis through the co-occurrence and clustering analysis applied by CiteSpace, to summarize the academic research keywords that can describe 'new power system ' keyword clusters of academic research, and finally, the keyword clusters are used as search fields for literature search and statistics, and the bibliometric method is used to analyze the research trend of 'new electric power system'.

2 Principle Analysis of CiteSpace's Co-occurrence and Clustering Functions

CiteSpace is an information visualization software developed in Java language, which is mainly based on the theory of co-citation analysis and path-finding network final algorithm[8], etc., to measure the literature (collection) of a specific field, in order to explore the key paths of the evolution of the subject area and its knowledge inflection points, and through a series of visualization mapping to form the analysis of the potential power mechanism of the evolution of the subject and the detection of the frontiers of the subject's development. This paper mainly uses CiteSpace. In this paper, we mainly use the text co-occurrence and clustering functions of CiteSpace to conduct bibliometric analysis.

Co-occurrence of scientific and technical texts is mainly a phenomenon in which knowledge units of the same or different types co-occur[9]. Examining the kinds of association within or between knowledge units belongs to the category of combinatorial mathematics. The knowledge unit of a scientific and technical text will contain n knowledge units such as title, author, organization, abstract, keywords, text, references, etc. The combination of these knowledge units (e.g., m knowledge paradigms are selected) can constitute a number of different types of co-occurrence analysis, the principle of which is as follows:

$$C_n^m = \frac{n!}{m!(n-m)!} \tag{1}$$

where m and n are both positive integers. When m=1, examine the co-occurrence of a particular knowledge module, such as the author co-occurrence analysis and

keyword co-occurrence analysis in this paper; when m>1, examine the connection between different different knowledge modules, which belong to the hybrid network, and are seldom applied in the actual analysis process.

In the co-occurrence analysis is divided into two categories: homo- and hetero-occurrence analysis, and both types of analysis can be expressed in mathematical expressions through the form of algebraic formulas and matrix diagrams. Let the heterocoincidence matrix formula for B_{ij} , meaning that when a knowledge module belongs to the text of the change, if it is recorded as 1, no is 0, the formula is:

$$B_{ij} = \begin{cases} 0, C_{ij} = 0 \\ 1, C_{ij} > 0 \end{cases} \tag{2}$$

And the same co-occurrence matrix, can also be expressed in mathematical formulas. Let A_{ij} matrix , we use 0 table of the studied knowledge module has no co-occurrence relationship, with n that the two have a co-occurrence relationship, and the knowledge module and co-occurrence of the number of times the more, the greater the value of n then the formula for:

$$A_{ij} = \begin{cases} 0, C_{ij} = 0 \\ n, C_{ij} = n, n \text{ is a positive integer} \end{cases} \tag{3}$$

C_{ij} in the formula is the number of occurrences of the ith modal knowledge block word and the jth knowledge module in the same paper. In which the homococcurrence matrix and heterococcurrence matrix are interconvertible as follows:

$$C\left[x_1, x_3\right] = O\left[x_1, x_2\right] \cdot O\left[x_2, x_3\right] \tag{4}$$

Clustering is the process of dividing a collection of physical or abstract objects into multiple classes[10], and the elements within the clustered subgroups have a high degree of similarity, while different subgroups have a high degree of dissimilarity. In this paper, we use the clustering algorithm of CiteSpace to divide the literature in the field of new power systems into several classes, and then analyze and summarize them to improve the keyword clusters.The clustering algorithm embedded in the CiteSpace software is based on the Modularity algorithm as the core.Modularity algorithm is intended to maximize the evaluation function (Q-function), and the bigger the Q-value is, the better the classification is, and the formula is as follows: the Q-value is the maximum, the better the classification is, the better the classification is. The formula is as follows:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - P_{ij}) \delta(C_i, C_j) \tag{5}$$

In the formula A_{ij} is the adjacency matrix of the network G, m is the total number of connections in the graph, P_{ij} is the expected value of the connection between node

i and node j, δ is the Kronecker function, C_i is the group to which the current node i belongs to, and the value of $\delta(C_i, C_j)$ is 1 when C_i and C_j belong to the same group, and 0 otherwise.

If the original and random graphs are assumed to have the same degree distribution, then $P_{ij} = \frac{k_i k_j}{2m}$, where k_i and k_j are the degrees of node i and node j, respectively, substituting P_{ij} into the above equation yields:

$$Q = \frac{1}{2m} \sum_{ij} (A_{ij} - \frac{k_i k_j}{2m}) \delta(C_i, C_j) \tag{6}$$

In the formula A_{ij} denotes the weight between node i and node j, k_i is the sum of the weights of the edges connected to node i, C_i is the grouping to which the current node i belongs, and the value of $\delta(C_i, C_j)$ is 1 when $C_i = C_j$, and 0 otherwise. if only one pair of nodes belongs to the same clustering, the above formula can be expressed as:

$$Q = \sum_{s=1}^{k} \left[\frac{m_s}{m} - \left(\frac{d_s}{2m} \right)^2 \right] \tag{7}$$

Where k denotes the number of clusters in the network, m denotes the total number of network connections, m_s denotes the number of connections contained in the cluster s, and d_s denotes the sum of node degrees in the cluster s of the network. The ideal Q value is between 0.3 and 0.7, and Q greater than 0.7 rarely occurs. The number of clusters is expressed by minimizing the network clustering objective function:

$$V\left(c_1 c_2 \cdots c_n\right) = \sum_{i<j} s_{ij} d_{ij}^2 - \sum_{i<j} d_{ij} \tag{8}$$

The $s_{ij} = \frac{2m}{c_i c_j}$ in the above formula can be converted into the maximization of the following formula by mathematical extrapolation to the above formula:

$$V\left(c_1 c_2 \cdots c_n\right) = \frac{1}{2m} \sum_{i<j} \delta(C_i, C_j) w_{ij} (c_{ij} - \gamma \frac{c_j c_j}{2m}) \tag{9}$$

Where $w_{ij} = \frac{2m}{c_i c_j}$, C_i is the cluster to which element i belongs, and $\delta(C_i, C_j)$ means that the equation value is 1 ($C_i = C_j$) or 0. γ is the resolution of the cluster, and the cluster distribution under different resolutions can be obtained by adjusting

the size of γ. The larger γ is, the more clusters are obtained and the more detailed the classification is.

$$\gamma = \begin{cases} 0, x_i = x_j \\ \dfrac{1}{d_{ij}}, x_i \neq x_j \end{cases} \tag{10}$$

When both w_{ij} and γ in the above formula are 1, it can be converted to:

$$V\left(c_1 c_2 \cdots c_n\right) = \frac{1}{2m}\sum_{i<j}(c_{ij} - \frac{c_j c_j}{2m})\delta(C_i, C_j) \tag{11}$$

Then simplify its original formula, the simplification process is as follows:

$$\begin{aligned} V\left(c_1 c_2 \cdots c_n\right) &= \frac{1}{2m}\sum_{i<j}\delta(C_i, C_j) w_{ij}(c_{ij} - \gamma\frac{c_j c_j}{2m}) \\ &= \frac{1}{2m}\sum_{i<j}\delta(C_i, C_j)\frac{2m}{c_j c_j}(c_{ij} - \gamma\frac{c_j c_j}{2m}) \\ &= \frac{1}{2m}\sum_{i<j}\delta(C_i, C_j)(s_{ij} - \gamma) \end{aligned} \tag{12}$$

It can be seen from this algorithm that the clustering of elements includes the parameters of clustering resolution, which makes it possible to increase or reduce the number of clusters according to the actual needs, and get more realistic results.

3 A Study of Keyword Clusters and Trends in Scientific and Technical Paper Distribution in the Context of New Power Systems

3.1 Definition of "New Power System" Literature Search Fields and Construction of Research Dataset

Based on the literature data retrieved by the "SU=('new power system' + 'carbon neutral' + 'new energy' + 'pumped storage' + 'energy storage technology' + 'virtual power plant' + 'low carbon' + 'demand response' + 'power market' + 'renewable energy') directive to construct the data set required for the study. After the subject and year are selected, the Refworks format literature data are exported through CNKI professional retrieval, and the website is shown in Fig. 1.

Fig. 1. Literature dataset required for the study

3.2 Constructing Novel Power System Keyword Clusters Based on Literature Visualization Analysis.

Analyze the keywords in CiteSpace, set the Node Type to the keyword Keywords, set the Time Slicing to 2018-2022 and Years Per Slice; the k-value of the g-index in the Selection Criteria is set to 50. run the CiteSpace software to derive the keyword network visualization knowledge graph of media convergence. Take the Top100 nodes as the object of analysis to get the merged network values of N=603, E=1255, i.e., the number of generated keyword nodes is 603, and the number of links is 1255. As shown in figure 2.

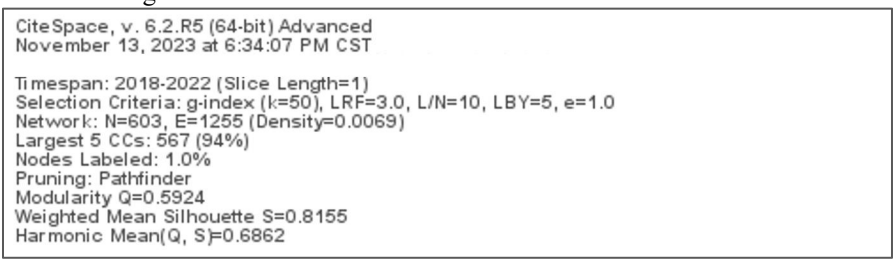

Fig. 2. Visualization of network operation-related parameters

Through the clustering analysis of the 2018-2022 novel power system literature data on the CiteSpace platform, the keywords appearing in the final novel power system research literature were classified into 14 major clustering centers, as shown in Figure 3, which are mainly focused on ‘virtual power plant’, ‘carbon neutral’, ‘demand response’, ‘power market’, etc. ‘Carbon Neutral’, ‘Demand Response’, ‘Electricity Market’ and so on.

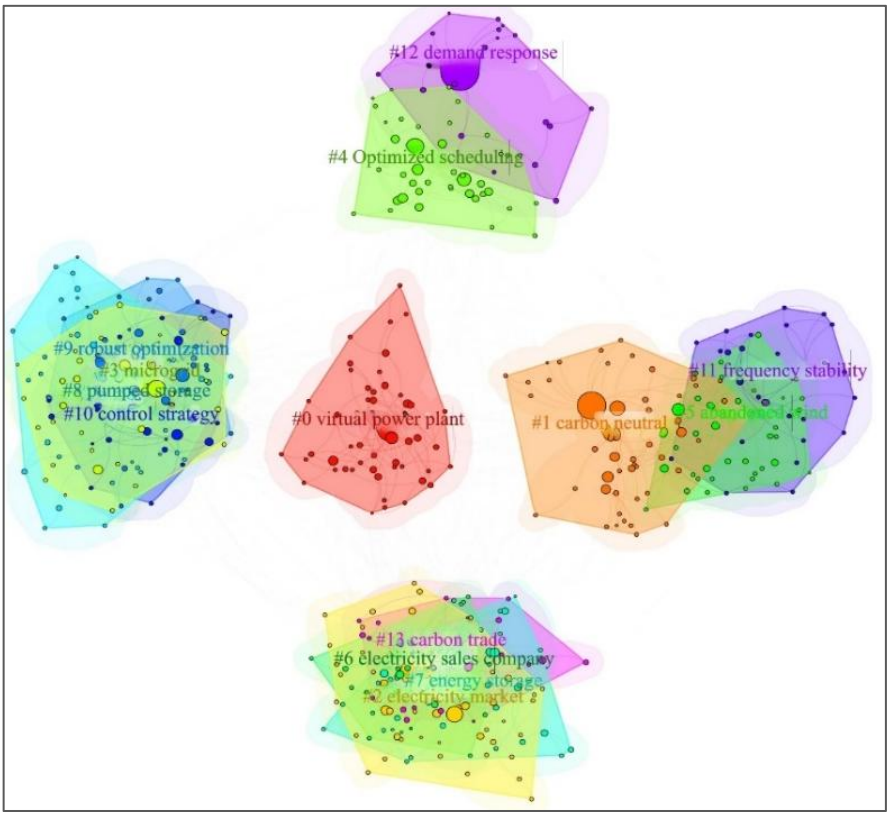

Fig. 3. Schematic of the 14 clustering centers

Based on the clarity of the network structure and clustering, CiteSpace provides two indicators, the module value (Q value) and the average profile value (S value), which can be used as a basis for us to judge the effectiveness of the mapping. Generally speaking, Q and S values are generally in the [0,1) interval, greater than 0.3 means that the delineation of the association structure is significant, when the S value of 0.7, the clustering is highly efficient and convincing, if more than 0.5, the clustering is generally considered reasonable. From the parameters related to the visualization network operation in Fig. 3: Q=0.5924, S=0.8155, the clustering result is reasonable.

Based on the above clustering results, excluding the keywords (such as "uncertainty") that are not clear enough in the clusters, and combining with the existing research reports, policies and experiences of experts in the field of new power system, we summarize seven keywords about "new power system". We summarize seven keyword clusters for the study of "new power system", taking into account the existing research reports and policies in the field and the experience of experts in the field. The keyword clusters are summarized in Table 1, as they highlight some of the research topics under the category of new power systems.

Table 1. Novel power system keyword clusters (>20 occurrences in the dataset)

No.	Topic	Keyword Cluster
1	Virtual Power Plant	virtual power plant, multi-complementary, optimized operation, data-driven
2	Carbon Neutral	new energy, carbon neutral, blockchain, peak carbon, carbon emission reduction, low carbon, energy transition, clean energy, carbon trading, wind power
3	Electricity Market	electricity market, spot market, ancillary services, trading mechanism, market mechanism, game theory
4	Smart Grid	electric vehicles, microgrids, flexible loads, load forecasting, smart grids, distribution grids, artificial intelligence, control strategies, coordinated control
5	Optimized Scheduling	Optimal dispatch, wind power consumption, master-slave game, two-tier optimization, economic dispatch, robust optimization
6	Energy Storage	energy storage, optimal allocation, peak shaving, pumped storage, energy storage systems
7	Demand Response	demand response, incentives, time-of-use tariffs, deep peaking, peaking, frequency stabilization

3.3 Constructed Keyword Clusters Based on the Study of the Output Volume and Publication Trends of Scientific Papers in the Context of New Power Systems

Based on the constructed keyword clusters, the research hotspots of new power systems in the past five years are studied. The keyword cluster of new power system is used as a CNKI database search field to analyze the trend of scientific and technological literature published in the research field of the keyword cluster from 2018 to 2022, and the bibliometric method is used to analyze the output of scientific and technological papers in the field of "new power system", and analyze the yearly trend of scientific and technological papers published under the themes, so as to provide a reference for the study of the development trend of the new power system. The bibliometric method is used to analyze the output of scientific research papers in the field of "new power system", and the annual trend of scientific papers published under each theme is analyzed, which provides a reference for the study of the development trend of new power system. The following figure shows the trend of scientific papers published in the field of "new power system":

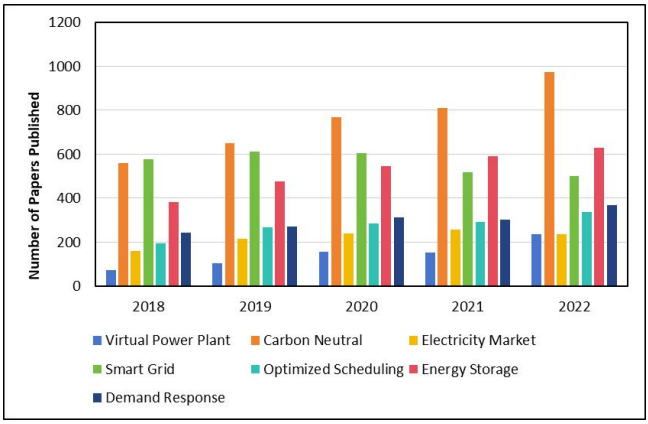

Fig. 4. Keyword Cluster Related Literature 2018-2022 Change in Number of Publications

As shown in Fig.4 , the six themes of 'Virtual Power Plant', 'Carbon Neutral', 'Electricity Market', 'Optimized Scheduling', 'Energy Storage' and 'Demand Response' show a general growth trend in 2018-2023; for the theme of 'Smart Grid', the trend of annual publications is relatively stable in 2018-2020, and the number of publications shows a more obvious downward trend in 2022-2021, while the number of publications tends to be stable in the following years.

4 Conclusions

The results of this paper show that for the CNKI database, there are seven main keyword clusters under the research category of 'new power system', namely: Virtual Power Plant, Carbon Neutral, Electricity Market, Smart Grid, Optimized Scheduling, Energy Storage, Demand Response, and each keyword cluster contains several keywords. Overall, the number of papers issued described by the seven keyword clusters shows a general upward trend during 2018-2022, indicating that research related to these seven areas will be a hotspot for the development of new power systems in the future.

Acknowledgment

This paper was supported by the self-funded project of Global Energy Interconnection Group Co. Ltd. Titled " Study on the clustering development strategy and data analysis system of sci-tech journals of State Grid" under grant 524500230002.

References

1. XU Xingxian.(2023).Research on network security protection of energy storage system of power plant.*Shandong Electric Power*, 50(6):6-12.

2. Ang, T. Z., Salem, M., Kamarol, M., Das, H. S., Nazari, M. A., & Prabaharan, N. (2022). A comprehensive study of renewable energy sources: Classifications, challenges and suggestions. *Energy Strategy Reviews*, *43*, 100939.
3. Diahovchenko, I., Kolcun, M., Čonka, Z., Savkiv, V., & Mykhailyshyn, R. (2020). Progress and challenges in smart grids: Distributed generation, smart metering, energy storage and smart loads. *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, *44*, 1319-1333.
4. LIANG Rongbo. (2023).ZHANG Yuntao,LIU Changfa,et al.Multi-source joint optimization scheduling to promote wind power consumption.*Shandong Electric Power*, 50(6):39-43.
5. Albert, M. J. (2021). The climate crisis, renewable energy, and the changing landscape of global energy politics. *Alternatives*, *46*(3), 89-98.
6. Kavic, M. S., & Satava, R. M. (2021). Scientific literature and evaluation metrics: impact factor, usage metrics, and altmetrics. *JSLS: Journal of the Society of Laparoscopic & Robotic Surgeons*, *25*(3).
7. Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*, *133*, 285-296.
8. Zhang, X., Zhang, Y., Wang, Y., & Fath, B. D. (2021). Research progress and hotspot analysis for reactive nitrogen flows in macroscopic systems based on a CiteSpace analysis. *Ecological Modelling*, *443*, 109456.
9. Nadi-Ravandi, S., & Batooli, Z. (2022). Gamification in education: A scientometric, content and co-occurrence analysis of systematic review and meta-analysis articles. *Education and Information Technologies*, *27*(7), 10207-10238.
10. Golalipour, K., Akbari, E., Hamidi, S. S., Lee, M., & Enayatifar, R. (2021). From clustering to clustering ensemble selection: A review. *Engineering Applications of Artificial Intelligence*, *104*, 104388.