Analysis Methods for Planning and Layout of Large Scientific Facility Clusters Based on Research Output Correlation

Authors: Feng Lingzi, Zhang Ruhao, Feng Kaiyue, Yuan Junpeng, Yuan Junpeng

Date: 2022-05-22T14:27:26Z

Abstract

[Purpose] Large scientific facilities play a crucial role in promoting the output of major scientific and technological achievements, and their clustered development and collaborative innovation effects continuously lead regional industrial innovation and development. To enable large facilities to better exert synergistic effects, it is necessary to improve the planning and layout of large facilities.

[Method] This study starts from scientific research outputs and proposes a general analytical framework that can be used to assist in analyzing the degree of discipline-level association between large scientific facilities, serving as a supplement to demand-driven planning. Taking the Guangdong Songshan Lake Science City as an example, this study collects paper data from existing large facilities (the Spallation Neutron Source and Synchrotron Radiation Light Source) and the planned large facility Y, as well as from similar types of facilities internationally, analyzes their intrinsic discipline-level associations, and uses this to assist in judging the matching and correlation degree between the planned facility and existing facilities.

[Results] The existing facilities (Spallation Neutron Source and Synchrotron Radiation Light Source) show high correlation and a potential complementary relationship, while the correlation between large facility Y and existing facilities is relatively low; if it is to be deployed, further expert demonstration is recommended.

[Limitations] This study primarily quantifies the connections between facilities based on their output achievements, and can only provide decision-makers with an objective data perspective and auxiliary reference. In terms of planning for large scientific facility clusters, national-level overall planning and regional demand for large scientific facilities are more important. This methodology

needs to be integrated with the intrinsic principles, operational processes of large scientific facilities, and expert knowledge to truly play a role in planning and layout.

[Conclusion] To a certain extent, this method can provide forward-looking quantitative references for the planning and layout of large facilities, thereby offering predictive and holistic optimization solutions for the functional coordination and future development of large facility clusters.

Full Text

Research on Analysis Methods for Large-Scale Scientific Facility Cluster Planning and Layout Based on Research Output Correlations

Feng Lingzi[1],[2], Zhang Ruhao[1],[2], Feng Kaiyue[3], Yuan Junpeng[1],[2]* [1] National Science Library, Chinese Academy of Sciences, Beijing 100190, China [2] Department of Library, Information and Archives Management, School of Economics and Management, University of Chinese Academy of Sciences, Beijing 100049, China [3] Condition Guarantee and Finance Bureau, Chinese Academy of Sciences, Beijing 100086, China

[**Objective**] Large-scale scientific facilities play a crucial role in promoting major scientific and technological achievements, and the collaborative innovation effects of their cluster development continuously drive regional industrial innovation. To maximize these synergistic effects, more effective planning of large-scale scientific facilities is essential. [**Methods**] This study proposes a general analytical framework based on scientific research outputs to assist in analyzing the degree of interdisciplinary correlation between large-scale scientific facilities, serving as a supplement to demand-driven planning. Using Guangdong's Songshan Lake Science City as a case study, we collected publication data from existing facilities (the Spallation Neutron Source and Synchrotron Radiation Source) and a planned facility Y from similar facilities worldwide, analyzing their intrinsic disciplinary-level correlations to evaluate the matching degree between the planned facility and existing ones. [**Results**] The analysis reveals high correlation between existing facilities (the Spallation Neutron Source and Synchrotron Radiation Source), suggesting potential complementarity. In contrast, facility Y shows relatively low correlation with existing facilities, indicating that further expert evaluation would be needed if this facility were to be included in the layout. [**Limitations**] This research primarily quantifies inter-facility connections based on research outputs, providing only an objective data perspective and auxiliary reference for decision-makers. For large-scale scientific facility cluster planning, national-level strategic planning and regional needs are more critical. This method must be integrated with the internal principles, operational processes, and expert knowledge of large-scale scientific facilities to play a substantive role in planning and layout. [**Conclusions**] This method can provide forward-looking quantitative references for facility planning and layout

to a certain extent, thereby offering a foreseeable and holistic optimization plan for functional coordination and future development of facility clusters.

Keywords: Large-scale scientific facility; Large-scale scientific facility cluster; Large-scale scientific facility layout; Major National Science and Technology Infrastructure; Scientific research output

---

1 Introduction

Large-scale scientific facilities are massive, complex research systems that provide extreme research capabilities for exploring the unknown world, discovering natural laws, and achieving technological transformations. They represent large-scale infrastructure essential for breaking through scientific frontiers and addressing major scientific and technological issues in economic development, social progress, and national security [1]. Currently, an increasing number of significant scientific discoveries both domestically and internationally rely on large-scale scientific facilities [3; 4]. Moreover, these facilities have demonstrated important and unique roles in integrating scientific and technological forces and resources, generating talent aggregation effects, and enhancing regional and national innovation capacity and international competitiveness [2]. Since 2009, China's Ministry of Science and Technology and the Chinese Academy of Sciences have proposed a series of plans, including the Medium- and Long-Term Plan for National Major Science and Technology Infrastructure Construction (2012-2030) and the 2050 Roadmap for Large-Scale Scientific Facility Development, which call for systematic deployment and construction of world-class large-scale scientific facilities oriented toward national strategic needs and frontier scientific development.

In recent years, the contradiction between the disciplinary limitations of individual large-scale scientific facilities and the increasingly interdisciplinary nature of scientific research has driven the cluster development of such facilities worldwide. Due to technical complementarity and other characteristics, the synergistic innovation effects among clustered facilities have become increasingly evident, continuously leading regional industrial innovation [2]. For example, Oak Ridge National Laboratory (ORNL) in the United States has assembled multiple large-scale scientific facilities, including the High Flux Isotope Reactor (HFIR), Oak Ridge Electron Linear Accelerator (ORELA), Spallation Neutron Source (SNS), and the "Summit" supercomputer, developing into a large comprehensive research base that has made tremendous contributions to neutron science, advanced materials, nanotechnology, and numerous other fields in the U.S. The Harwell Science and Innovation Campus (HSIC) in the UK has brought together major scientific infrastructure such as the Diamond synchrotron light source, ISIS pulsed spallation neutron source, Central Laser Facility (CLF), and New Light Source (NLS), attracting a large number of world-leading scientists in physics, chemistry, materials science, geology, biomedicine, and environmental

science, and becoming a world-leading high-tech park integrating science, innovation, and enterprise. The High Energy Accelerator Research Organization (KEK) in Japan has assembled facilities including the KENS pulsed spallation neutron facility, PF-AR photon factory advanced ring, and J-PARC joint proton synchrotron accelerator, playing an important role in enhancing Japan's scientific and technological competitiveness. Through joint construction by central and local governments, China has mobilized local enthusiasm and successively deployed a batch of large-scale scientific facilities in Beijing, Hefei, Shanghai, and other locations, forming clusters of large-scale scientific facilities.

The significance of large-scale scientific facility clusters is substantial. They address key and common problems in industrial technology development by forming knowledge synergy through facility and knowledge information sharing, achieving interdisciplinary integration and complementary advantages, thereby breaking through technical barriers and achieving original knowledge innovation at more complex and microscopic levels [2]. Furthermore, the "agglomeration effect" and "benchmarking effect" of talent, resources, cooperation, and industry brought by large-scale scientific facility clusters will better promote large-scale, wide-ranging regional economic development and industrial upgrading. The potential benefits of large-scale scientific facility clusters are particularly necessary for China, which is in a critical period of building a strong science and technology nation. The 14th Five-Year Plan explicitly states that regional systematic layout of national major science and technology infrastructure should be carried out from four functional dimensions: strategic orientation, application support, forward-looking guidance, and livelihood improvement, thereby supporting Beijing, Shanghai, and the Guangdong-Hong Kong-Macao Greater Bay Area in forming international science and technology innovation centers, and establishing comprehensive national science centers in Beijing Huairou, Shanghai Zhangjiang, the Greater Bay Area, and Hefei, Anhui [6].

Large-scale scientific facilities involve enormous investment. To maximize their practical value, rational layout is required under the premise of conforming to national overall planning and regional demand for such facilities. Research on the layout of large-scale scientific facility clusters has gradually attracted attention from domestic academia. Some scholars view large-scale scientific facilities as "high-density aggregations of scientific and technological innovation resources" and as powerful support for the Guangdong-Hong Kong-Macao Greater Bay Area to cope with global competition for innovation resources. However, they also note that the layout of large-scale scientific facilities in Guangdong Province and even nationwide suffers from excessive dispersion, requiring greater aggregation and more rational layout to enhance the effectiveness of innovation resource concentration [7]. Other scholars have investigated international innovation parks and bay areas that rely on large-scale scientific facilities, examining the current status and development trends of such facilities to draw lessons for cluster layout. For instance, Zhang Lingling et al. discussed the spatial layout and functional positioning of the spallation neutron source large-scale scientific facility in science parks [8]; Chen Tao and Feng Feng emphasized the conno-

tation, advantages, and cluster effects of large-scale scientific facility clusters, calling for cluster development to stimulate innovation and interdisciplinary potential [5]; Chen Anming and Wei Dongyuan identified shortcomings in the overall planning, talent supply, and cluster planning of major science and technology infrastructure in the Guangdong-Hong Kong-Macao Greater Bay Area through comparative analysis [9]. Cheng Xiaofang et al. selected the Yangtze River Delta as a typical region, using economic modeling and deductive methods to analyze optimization ideas and social benefits of large-scale scientific facility layout. Their research found that rational application and functional planning, strengthened joint construction, management, and sharing of facility clusters, and promotion of regional science-technology-economic linkage development would help truly realize the role of large-scale scientific facilities [10].

Some scholars also believe that disciplinary layout structure affects the scientific benefits of large-scale scientific facilities [11]. For example, Zhang Lingling et al. analyzed the disciplinary themes and cooperation networks of the Shanghai Synchrotron Radiation Facility as a case study [11]. In contrast, foreign scholars have paid less attention to the preliminary layout research of large-scale scientific facilities and clusters, focusing mainly on various benefits such as economic and social benefits [12; 13], industrial innovation efficiency [14], research output performance (e.g., number of papers [15; 16], impact factors [17]), and talent cultivation and cooperation [18].

Overall, although current research has focused on the planning, layout, and potential benefits of large-scale scientific facility clusters, most remain at the theoretical discussion level without addressing practical layout strategies. While acknowledging the potential for interdisciplinary integration and scientific innovation that cluster layout may stimulate, few studies have specifically analyzed layout rationality from the perspective of disciplinary fit within large-scale scientific facility clusters. Existing large-scale scientific facility clusters demonstrate disciplinary agglomeration effects, and under the current situation where interdisciplinary research greatly promotes innovation, rational disciplinary planning can further enhance collaborative effects and stimulate innovation. Therefore, systematic disciplinary layout in the planning and construction process of large-scale scientific facilities is an urgent practical problem [11]. Currently, the actual layout work of large-scale scientific facility clusters is primarily driven by macro-level national needs [5]. In preliminary project demonstrations, there exists a situation of "seeing only the trees but not the forest" and "seeing only disciplines but not their interconnections," with insufficient demonstration of relationships between carriers and between carriers and projects, and inadequate reflection of frontier guidance and comprehensive interdisciplinary nature. Therefore, relevant research urgently needs to be conducted to strengthen the study of disciplinary correlations between facilities and between facilities and platforms, clarify cross-field layout under major disciplinary planning, and form disciplinary positioning characteristics in pilot zones.

The important contribution of this paper is to provide quantitative decision-

making references for disciplinary layout to supplement demand-driven facility planning. A more closely connected and rational layout of large-scale scientific facilities can improve benefits and generate more significant and timely positive impacts on the national economy. To provide decision-making reference information for more scientific layout of large-scale scientific facility clusters at the disciplinary level, this study proposes a general analytical framework based on scientific research outputs to assist in analyzing the degree of disciplinary-level correlation between facilities. This framework helps provide forward-looking quantitative references for facility planning and layout, thereby offering optimization solutions for functional coordination and future development of facility clusters from a disciplinary layout perspective, serving as a supplement to demand-driven planning.

It should be noted that while this research method has certain reference significance for the planning of large-scale scientific facility clusters, the more important factors in cluster planning are national-level strategic planning and regional demand for such facilities.

2 Analytical Framework for Large-Scale Scientific Facility Cluster Planning and Layout

Since disciplinary layout structure significantly impacts the benefits generated by large-scale scientific facilities, enabling clusters to maximize their effectiveness through inter-facility synergy requires sufficient intrinsic correlation at both the disciplinary and research levels. Therefore, we can characterize the disciplinary and research-level correlations among facilities by obtaining their publication output data.

Figure 1. Technical Route

The technical route of our methodology is illustrated in Figure 1, with specific steps as follows:

(1) Confirm Analysis Objects: Identify the target large-scale scientific facilities for analysis. To analyze correlations among facilities in an existing cluster, we can use their published research outputs for characterization. To analyze the correlation degree between a planned facility and existing ones, we can use publication data from similar facilities that have already been deployed as a proxy.

(2) Acquire Facility Data: Obtain publication data from various large-scale scientific facilities through their official publication libraries, annual reports, impact assessment reports, official news releases, and consultation with facility publication library administrators, using methods including direct download, email requests, and web scraping.

(3) Extract Paper Information: Retrieve bibliographic information from paper databases (such as Web of Science) using paper identifiers like DOI, and parse disciplinary classifications and citation relationships.

(4) Characterize Facility Knowledge Distribution: Determine the knowledge distribution of facilities based on the disciplinary classifications of their publications. Since papers are produced through research conducted at large-scale scientific facilities and belong to specific disciplines, we can derive the correspondence between facilities and disciplines, thereby obtaining the knowledge distribution of facilities across various disciplines.

Construct a "facility-discipline" correspondence to identify the major disciplines involved in the analyzed facilities and their cumulative paper proportions, as shown in Figure 2. Darker lines from a facility to a discipline indicate higher proportions of papers in that discipline; darker nodes on the right indicate disciplines that serve as major knowledge sources for multiple facilities.

This study employs both broad disciplines and sub-disciplines in its analysis. Broad disciplines use the ESI classification system[1], while sub-disciplines use the WoS classification system[2].

Figure 2. Schematic Diagram of Facility-Discipline Knowledge Proportion

[1] ESI (Essential Science Indicators) is a bibliometric analysis database built upon SCI (Science Citation Index) and SSCI (Social Sciences Citation Index) records, comprising 22 disciplinary categories.

[2] Web of Science is Clarivate Analytics' global academic database platform, including SCI, SSCI, A&HCI, and other databases, covering 252 disciplinary categories.

(5) Facility Correlation Analysis: This study achieves "facility-to-facility" correlation analysis through two complementary pathways:

Pathway 1: Construct a "facility-discipline" bipartite network to derive pairwise facility correlations through disciplinary co-occurrence and distribution similarity.

Figure 3. Facility-Discipline Network Schematic

As shown in Figure 3, we construct a two-layer network with a facility layer and a discipline layer. Since papers are produced through research at large-scale scientific facilities and belong to specific disciplines, we can derive facility-discipline correspondences and consequently inter-facility disciplinary co-occurrence connections. However, conventional disciplinary co-occurrence strength algorithms may be influenced by different publication volumes across facilities. To more accurately measure the similarity of disciplinary distributions among facilities, we employ cosine similarity of disciplinary distribution vectors, defined as:

$$S(A, B) = \theta(v_A, v_B) = \frac{v_A \cdot v_B}{\|v_A\|\|v_B\|}$$

where $\Phi = [\varphi_1, \varphi_2, ..., \varphi_m]$ represents all existing sub-disciplinary classifications,

and v_A, v_B represent the disciplinary distribution vectors of facilities A and B, respectively. For example, $v_A = [v_A^{\varphi_1}, v_A^{\varphi_2}, ..., v_A^{\varphi_m}]$, where $v_A^{\varphi_1}$ denotes the proportion of papers from facility A in discipline φ_1 relative to A' s total publications.

Pathway 2: Construct a "facility-paper" bipartite network to indirectly derive facility correlations through paper-level relationships.

Figure 4. Facility-Paper Network Schematic

As shown in Figure 4, we construct a two-layer network with a facility layer and a paper layer. Papers published based on facility research may have citation relationships, allowing facility correlations to be indirectly characterized through paper-level connections. For instance, if paper 1 cites paper 4, and paper 2 cites paper 3, with papers 1 and 2 based on facility A and papers 3 and 4 based on facility B, then a raw weighted edge from facility A to facility B with weight 2 is formed in the facility layer, and so on.

Let P_A, P_B be the sets of papers published based on facilities A and B, and R_A, R_B be the reference sets of P_A, P_B. The direct citation relationship strength $DCS(A, B)$ and bibliographic coupling strength $BCS(A, B)$ between facilities A and B can be defined as:

$$DCS(A, B) = AVG \left(\frac{|R_A \cap P_B|}{|P_B|}, \frac{|R_B \cap P_A|}{|P_A|} \right)$$

$$BCS(A, B) = \frac{|R_A \cap R_B|}{|R_A \cup R_B|}$$

(6) Results Interpretation: Analyze and interpret the correlation patterns among large-scale scientific facilities based on the data results.

3 Case Study and Data Acquisition

This study uses Guangdong' s Songshan Lake Science City as a case to analyze correlations between existing and planned large-scale scientific facilities. Songshan Lake Science City has already deployed the China Spallation Neutron Source, with the Southern Advanced Light Source research platform and pre-research projects actively underway. The science city is showing initial trends of forming a major science and technology infrastructure cluster, with gradually emerging agglomeration effects on high-end innovation resources. Based on this foundation, a new large-scale facility X is planned for addition. To maximize synergistic effects, the newly planned facility should demonstrate strong disciplinary-level correlation with existing facilities to ensure inter-facility synergy.

(1) Identify Target Facilities for Analysis: Since the number of publications from existing facilities at Songshan Lake Science City is currently limited,

this study selected similar facilities domestically and internationally as proxies, obtaining their published paper data for analysis. The selected proxy facilities are shown in Table 1.

For spallation neutron source facilities, we selected three internationally established facilities as proxies: the U.S. Spallation Neutron Source (SNS), UK Spallation Neutron Source (ISIS), and Japanese Spallation Neutron Source (J-PARC).

For the Southern Light Source proxy selection, we first compiled a list of established light sources at similar energy levels as candidates. The Shanghai Synchrotron Radiation Facility, as China's successful third-generation synchrotron light source, would have been ideal as both a domestic and same-energy-level facility. However, data acquisition channels and the final data volume were extremely limited, with only 127 reported achievements obtained, making direct analysis significantly biased compared to other facilities. Considering data completeness and accessibility, we ultimately selected the UK Diamond Light Source and Spanish ALBA Light Source for analysis.

For the planned large-scale facility X, we selected a similar international facility Y as the proxy.

Table 1. Target Facilities at Songshan Lake Science City and Proxy Facilities

Songshan Lake Science City Facilities	Proxy Facilities
Spallation Neutron Source	U.S. Spallation Neutron Source (SNS), UK Spallation Neutron Source (ISIS), Japanese Spallation Neutron Source (J-PARC)
Southern Light Source	UK Diamond Light Source, Spanish ALBA Light Source
Planned Facility X	International Facility Y

(2) Data Acquisition: After confirming the target facility list, we obtained publication data from each facility through official publication libraries, news releases, annual reports, impact assessment reports, and consultation with facility publication library administrators, using direct downloads, email requests, and web scraping. Using the obtained DOIs, we matched and batch-retrieved bibliographic data from the Web of Science database. All data were collected in August 2020.

4.1 Publication Statistics

Publications based on spallation neutron sources are shown in Figure 5. As of August 2020, a total of 14,384 papers were published, including 11,381 from the UK facility, 2,100 from the U.S. facility, and 903 from the Japanese facility. Publication numbers have fluctuated and increased since 1978, declined between 2004-2009, and showed significant growth from 2009 onward, peaking at 907 papers in 2018.

Figure 5. Publication Statistics for Spallation Neutron Sources

Publications based on synchrotron radiation light sources are shown in Figure 6. As of August 2020, a total of 10,421 papers were published, including 9,141 from the UK Diamond Light Source and 1,280 from the Spanish ALBA Light Source. Publication numbers have grown nearly exponentially since 2001, with growth slowing around 2017 and peaking at 1,398 papers in 2019.

Figure 6. Publication Statistics for Synchrotron Radiation Light Sources

Publications from facility Y (similar to the planned facility X) are shown in Figure 7. Due to its relatively recent completion, only 247 papers were published as of August 2020. Publication numbers have continuously increased since 2013, peaking at 66 papers in 2018 before declining.

Figure 7. Publication Statistics for Facility Y

4.2 Disciplinary Distribution

To understand the disciplinary attributes of the three types of scientific facilities at Songshan Lake Science City, we first conducted disciplinary analysis of their international counterparts (including U.S. SNS, UK ISIS, Japanese J-PARC, Spanish ALBA, UK Diamond, and facility Y).

At the broad discipline level (ESI classification), physics and chemistry are the two major disciplines most commonly addressed by all three facility types, accounting for 66.7%, 41.8%, and 78.9% of relevant papers respectively, demonstrating the irreplaceable role of large-scale scientific facilities in basic disciplinary research. Materials science and biochemistry/molecular biology also constitute significant proportions. Additionally, all three facility types include some interdisciplinary research, indicating that beyond supporting multiple major disciplines, these facilities also possess the potential to support interdisciplinary scientific research.

Table 2. Top 5 Broad Discipline Distributions for Three Facility Types

Rank	Spallation Neutron Source	Synchrotron Light Source	Facility Y
1	Physics (5,324, 37.0%)	Physics (4,275, 29.7%)	Physics (168, 68%)
2	Chemistry (3,070, 29.5%)	Chemistry (1,286, 12.3%)	Chemistry (27, 10.9%)
3	Materials Science (1,681, 11.7%)	Biochemistry/Molecular Biology (1,278, 12.3%)	Materials Science (15, 6.1%)
4	Biochemistry/Molecular Biology (1,267, 12.2%)	Materials Science (1,073, 10.3%)	Biochemistry/Molecular Biology (14, 5.7%)
5	Interdisciplinary Sciences (415, 2.9%)	Interdisciplinary Sciences (168, 1.6%)	Interdisciplinary Sciences (6, 2.4%)

Examining interdisciplinary patterns through the distribution of disciplines, synchrotron light sources demonstrate a relatively balanced interdisciplinary structure, with chemistry, physics, interdisciplinary sciences, biochemistry/molecular biology, and materials science each accounting for over 10%. In contrast, facility Y shows the weakest interdisciplinary balance, with physics alone accounting for 68% of publications and other disciplines scattered. This phenomenon may arise from two reasons: First, facility Y currently serves primarily as a means to study material properties and discover new phenomena in optical physics, condensed matter physics, and nuclear physics, while spallation neutron sources and synchrotron light sources have broader applications. Second, facility Y represents a recent breakthrough in laser technology and ultrafast science this century, with few mature facilities worldwide, and facility immaturity has led to disciplinary structure skewness.

At the sub-discipline level (WoS classification), fine-grained analysis of relevant sub-disciplines for large-scale scientific facilities better reveals inter-facility disciplinary correlations. Among the obtained publication data, spallation neutron sources involve 108 sub-disciplines, synchrotron light sources involve 131, and facility Y involves 38. The top 10 sub-disciplines for each facility show that Materials Science-Multidisciplinary, Condensed Matter Physics, and Chemical Physics are the most relevant for spallation neutron sources; Biochemistry/Molecular Biology, Materials Science-Multidisciplinary, and Chemistry-Multidisciplinary are most relevant for synchrotron light sources; and Optics, Atomic/Molecular/Chemical Physics are most relevant for facility Y. The three facility types show dense overlap in relevant sub-disciplines.

Table 3. Top 10 Sub-Discipline Distributions for Three Facility Types

Rank	Spallation Neutron Source	Synchrotron Light Source	Facility Y
1	Materials Science, Multidisciplinary	Biochemistry & Molecular Biology	Optics
2	Condensed Matter Physics	Materials Science, Multidisciplinary	Atomic, Molecular, Chemical Physics
3	Chemical Physics	Chemistry, Multidisciplinary	Physics, Multidisciplinary
4	Applied Physics	Chemical Physics	Instruments & Instrumentation
5	Chemistry, Multidisciplinary	Applied Physics	Nuclear Science & Technology
6	Physics, Multidisciplinary	Condensed Matter Physics	Fluid & Plasma Physics
7	Fluid & Plasma Physics	Biophysics	Nanoscience & Nanotechnology
8	Instruments & Instrumentation	Cell Biology	Materials Science, Multidisciplinary
9	Biochemistry & Molecular Biology	Fluid & Plasma Physics	Applied Physics
10	Nuclear Science & Technology	Nanoscience & Nanotechnology	Chemical Physics

To visually represent the disciplinary knowledge distribution of the three facility types, we constructed facility-sub-discipline distribution relationships based on sub-discipline distributions (Figure 8). In this figure, darker lines indicate higher proportions of a sub-discipline in a facility' s publications; darker nodes on the right represent sub-disciplines that carry more disciplinary weight from multiple facilities.

Figure 8. Facility-Sub-Discipline Distribution Relationships

Figure 8 shows that Materials Science-Multidisciplinary, Chemical Physics, and Applied Physics are the most overlapping sub-disciplines among the three facility types. Additionally, Condensed Matter Physics, Chemistry-Multidisciplinary, and Crystallography show dense overlap between spallation neutron sources and synchrotron light sources, while Atomic/Molecular/Chemical Physics, Physics-

Multidisciplinary, and Instruments & Instrumentation show considerable overlap between spallation neutron sources and facility Y. These disciplines require special attention when supporting cluster development, and relevant research strengths in these areas could be considered as supporting research forces.

4.3 Facility Correlation Analysis

To more quantitatively explore similarities and differences among the three facility types, we conducted analyses based on both disciplinary distribution vectors and paper-level networks.

Regarding disciplinary distribution vectors, Table 4 presents pairwise cosine similarity comparisons of disciplinary distributions. The similarity between spallation neutron sources and synchrotron light sources reaches 0.738. Combined with Table 2, we observe that synchrotron light sources involve Biochemistry & Molecular Biology (11.32%), Biophysics (4.39%), and Cell Biology (3.17%) —disciplines not primarily addressed by spallation neutron sources. This indicates that spallation neutron sources and synchrotron light sources have both moderate disciplinary distribution similarity and distinct primary disciplinary domains, making their co-layout reasonable from a research output perspective. Facility Y shows medium-to-low similarity with the other two (0.469 and 0.406), but possesses major relevant disciplines not found in the others—Optics (22.05%), Fluid & Plasma Physics (5.57%), Nuclear Science & Technology (3.79%), and Nanoscience & Nanotechnology (3.34%)—suggesting that facility Y could provide certain disciplinary complementarity when co-located with spallation neutron sources and synchrotron light sources.

Table 4. Disciplinary Distribution Similarities Among Three Facility Types

	Spallation Neutron Source	Synchrotron Light Source	Facility Y
Spallation Neutron Source	1.000	0.738	0.469
Synchrotron Light Source	0.738	1.000	0.406
Facility Y	0.469	0.406	1.000

Through cluster distributions in direct citation and bibliographic coupling networks, we can visually demonstrate the correlation of research topics among papers from different facilities, thereby indirectly revealing inter-facility connections. Figures 9 and 10 present citation network analyses of publications from the three facility types. Figure 9 reveals network relationships arising from mutual citations, while Figure 10 shows relationships from reference coupling. Both figures yield similar conclusions: First, spallation neutron source papers

(orange nodes) and synchrotron light source papers (purple nodes) form densely connected clusters, while synchrotron light sources also have an independent cluster, indicating strong association and synchronization in some research areas, with synchrotron light sources covering larger fields such as biochemistry and molecular biology that are less addressed by spallation neutron sources. Second, facility Y papers (green nodes) occupy peripheral positions in the network diagrams, related to the scarcity of mature facilities worldwide and limited publications. Based on current publications, facility Y shows relatively weak connections with the other two facility types.

Figure 9. Facility-Paper Direct Citation Hierarchical Network

Figure 10. Facility-Paper Bibliographic Coupling Hierarchical Network

These quantitative data lead to the following conclusions: All three facility types provide major support for basic disciplines and possess varying degrees of potential for interdisciplinary research. They share common or similar relevant sub-disciplines, including Chemical Physics, Applied Physics, Condensed Matter Physics, Crystallography, Atomic/Molecular/Chemical Physics, Instruments & Instrumentation, and some interdisciplinary sciences, requiring special attention in future co-layout planning. Research strengths in these disciplines could serve as supporting forces. Spallation neutron sources and synchrotron light sources show strong disciplinary distribution and research topic similarity, with synchrotron light sources providing clear disciplinary and research topic complementarity to spallation neutron sources. Facility Y currently has relatively low disciplinary correlation and research topic similarity with the other two, but plays a non-negligible role in optics, atomic/molecular/chemical physics.

Overall, at the data level, spallation neutron sources and synchrotron light sources show high correlation with potential complementarity, while facility Y shows lower correlation with the other two, suggesting that further expert evaluation would be needed if layout is proposed.

5 Limitations and Future Work

This study proposes a general quantitative analytical framework to assist in planning and layout of large-scale scientific facility clusters. Using Guangdong's Songshan Lake Science City as a case study, we collected publication data from similar facilities worldwide for both existing and planned facilities, analyzing their intrinsic disciplinary-level correlations to evaluate matching degrees.

This research has several limitations: (1) The method currently only interprets data quantitatively to explore disciplinary correlations without qualitative analysis, which may introduce certain accuracy limitations. (2) The analysis only addresses matching degrees from a disciplinary correlation perspective, while actual facility matching involves additional dimensions (research level, geography, economic/industrial level, talent level, etc.), with disciplinary relationships

being just one important aspect. (3) The study only uses facility-based publications as analysis data, which should be expanded for more comprehensive results. (4) The analysis only uses existing data, presenting a summary of the past, while facilities have development potential that may generate new uses as clusters expand. (5) For planned facilities or recently deployed facilities with limited outputs, the analysis uses worldwide similar facilities as proxies, which may deviate from the actual analysis object due to subtle differences in energy levels, cluster configurations, and other factors. Additionally, due to varying data accessibility, more suitable proxy facilities were omitted in the light source analysis, potentially affecting accuracy.

Future work will further refine the method, explore more quantifiable elements in facility cluster relationships, expand analysis data, and combine qualitative and quantitative analyses to make results more valuable for facility planning and layout, thereby providing foreseeable and holistic optimization for functional coordination and future development of facility clusters.

It should be emphasized that, like any methodological system, scientometric methods have limitations and cannot cover everything. This study primarily quantifies inter-facility connections from the perspective of facility publications, providing only an objective data perspective and auxiliary reference for decision-makers. For large-scale scientific facility cluster planning, national-level strategic planning and regional demand are more critical. This method must be integrated with the internal principles, operational processes, and expert knowledge of large-scale scientific facilities to play a substantive role in planning and layout.

References

[1] Wang Jichang. Modern Science and Technology Terminology Selection [M]. Zhengzhou: Henan Science and Technology Press, 2006.

[2] Liang Yongfu, Pan Sitao, Lin Xiong, et al. Collaborative Innovation and Industrial Driving Effects of Large-Scale Scientific Facility Clusters—A Case Study of Guangdong Large Science Center [J]. Science and Technology Management Research, 2018, 38(03): 5-10.

[3] Xiao Guoqing, Li Zhenzhong. Lanzhou Heavy Ion Research Facility [J]. Bulletin of Chinese Academy of Sciences, 2009, 24(01): 97-101+2+105.

[4] Qi Fang: National Heavy Instruments Lay Foundation for Innovative Future —A Review of China's Large-Scale Scientific Facility Achievements Since the 18th National Congress [N], Guangming Daily.

[5] Chen Tao, Feng Feng. Cluster Effects and Management Implications of Large-Scale Scientific Facilities [J]. Journal of Northwestern Polytechnical University (Social Sciences Edition), 2015, 35(01): 61-66.

[6] The 14th Five-Year Plan for National Economic and Social Development

and Long-Range Objectives Through 2035 [EB/OL]. (2021-08-24) [2021-08-24]. http://www.gov.cn/xinwen/2021-03/13/content_5592681.htm.

[7] Huang Zhenyu. Analysis of the Relationship Between Large-Scale Scientific Facilities and Guangdong-Hong Kong-Macao Greater Bay Area Development [J]. Science and Technology Management Research, 2019, 39(18).

[8] Zhang Lingling, Zhao Minghui, Zhao Daozhen, et al. Research on Spatial Layout and Countermeasures of Science Parks Based on Large-Scale Scientific Facilities—A Case Study of Spallation Neutron Source [J]. Engineering Studies, 2019, 11(04): 338-348.

[9] Chen Anming, Wei Dongyuan. Optimization Analysis of Major Science and Technology Infrastructure Layout in Guangdong-Hong Kong-Macao Greater Bay Area—Based on International Comparison [J]. International Economics and Trade Research, 2020, 36(10): 86-99.

[10] Cheng Xiaofang, Tang Lei, Xia Yilin. Co-construction and Sharing of Large-Scale Scientific Facilities and Their Impact on Regional Integration—A Case Study of Yangtze River Delta [J]. Science and Technology Management Research, 2020, 40(22): 26-31.

[11] Zhang Lingling, Zhao Minghui, Zeng Gang, et al. Research on Disciplinary Themes and Cooperation Networks Based on Large-Scale Scientific Facilities from a Bibliometric Perspective—A Case Study of Shanghai Synchrotron Radiation Facility [J]. Management Review, 2019, 31(11): 279-288.

[12] Florio M, Sirtori E. Social benefits and costs of large scale research infrastructures[J]. Technological Forecasting and Social Change, 2016, 112: 65-78.

[13] Florio M, Forte S, Sirtori E. Forecasting the socio-economic impact of the Large Hadron Collider: A cost-benefit analysis to 2025 and beyond[J]. Technological Forecasting and Social Change, 2016, 112: 38-53.

[14] Michalowski S. The Impacts of Large Research Infrastructures on Economic Innovation and on Society: Case Studies at CERN[R]. Europe: Organisation for Economic Co-operation and Development (OECD), 2014.

[15] Hallonsten O. How expensive is Big Science? Consequences of using simple publication counts in performance assessment of large scientific facilities[J]. Scientometrics, 2014, 100(2).

[16] Hallonsten O. Use and productivity of contemporary, multidisciplinary Big Science[J]. Research Evaluation, 2016, 25(4): 486-495.

[17] Heidler R, Hallonsten O. Qualifying the performance evaluation of Big Science beyond productivity, impact and costs[J]. Scientometrics, 2015, 104(1): 295-312.

[18] Qiao L L, Mu R P, Chen K H. Scientific effects of large research infrastructures in China[J]. Technological Forecasting and Social Change, 2016, 112: 102-112.

Author Contributions

Feng Lingzi: Conceived research ideas, designed research framework, collected and analyzed data, wrote and revised the paper.

Zhang Ruhao: Designed research framework, collected, cleaned, and analyzed data, wrote and revised the paper.

Feng Kaiyue: Proposed research propositions, analyzed case selection, revised the paper.

Yuan Junpeng: Conceived research ideas, revised the paper.

Note: Figure translations are in progress. See original paper for figures.

Source: ChinaXiv — Machine translation. Verify with original.