

Research evolution of metal organic frameworks: A scientometric approach with human-in-the-loop

Citation: Zhao, X. T., Langlois, K., Furst, J., An, Y., Hu, X. H., Gualdrón, D. G., Uribe-Romo, F., & Greenberg, J. (2024). Research evolution of metal organic frameworks: A scientometric approach with human-in-the-loop. *Journal of Data and Information Science*, 9(3), 44–64. <https://doi.org/10.2478/jdis-2024-0019>

Received: Mar. 19, 2024

Revised: May 01, 2024;
Jun. 18, 2024

Accepted: Jun. 19, 2024

Xintong Zhao^{1†}, Kyle Langlois², Jacob Furst², Yuan An¹, Xiaohua Hu¹, Diego Gomez Gualdrón³, Fernando Uribe-Romo², Jane Greenberg¹

¹Drexel University, 3141 Chestnut St, Philadelphia, PA, USA

²University of Central Florida, 4000 Central Florida Blvd, Orlando, FL, USA

³Colorado School of Mines, 1500 Illinois St, Golden, CO, USA

Abstract

Purpose: This paper reports on a scientometric analysis bolstered by human-in-the-loop, domain experts, to examine the field of metal-organic frameworks (MOFs) research. Scientometric analyses reveal the intellectual landscape of a field. The study engaged MOF scientists in the design and review of our research workflow. MOF materials are an essential component in next-generation renewable energy storage and biomedical technologies. The research approach demonstrates how engaging experts, via human-in-the-loop processes, can help develop a comprehensive view of a field's research trends, influential works, and specialized topics.

Design/methodology/approach: A scientometric analysis was conducted, integrating natural language processing (NLP), topic modeling, and network analysis methods. The analytical approach was enhanced through a human-in-the-loop iterative process involving MOF research scientists at selected intervals. MOF researcher feedback was incorporated into our method. The data sample included 65,209 MOF research articles. Python3 and software tool *VOSviewer* were used to perform the analysis.

Findings: The findings demonstrate the value of including domain experts in research workflows, refinement, and interpretation of results. At each stage of the analysis, the MOF researchers contributed to interpreting the results and method refinements targeting our focus



on MOF research. This study identified influential works and their themes. Our findings also underscore four main MOF research directions and applications.

Research limitations: This study is limited by the sample (articles identified and referenced by the Cambridge Structural Database) that informed our analysis.

Practical implications: Our findings contribute to addressing the current gap in fully mapping out the comprehensive landscape of MOF research. Additionally, the results will help domain scientists target future research directions.

Originality/value: To the best of our knowledge, the number of publications collected for analysis exceeds those of previous studies. This enabled us to explore a more extensive body of MOF research compared to previous studies. Another contribution of our work is the iterative engagement of domain scientists, who brought in-depth, expert interpretation to the data analysis, helping hone the study.

Keywords Scientometric; Metal-Organic Frameworks (MOFs); Network analysis; Topic modeling; Human-in-the-loop

1 Introduction

Recent advancements in environmental science, biomedicine, and computing techniques have motivated development of an evolving, cross-disciplinary field known as “metal-organic framework (MOF)”, which has garnered significant interest. Due to their unique material characteristics, MOF materials are often considered as a platform for various next-generation applications, such as renewable energy storage, carbon capture, drug delivery and disease diagnosis, among other applications (Cavka et al., 2008; Cote et al., 2005; Wang et al., 2016).

As an emerging field, MOF research integrates a range of chemical synthesis disciplines (Ginsberg, 1990; Kinoshita, Matsubara, & Saito, 1959; Moulton & Zaworotko, 2001) into what is known as Reticular Chemistry (Yaghi, Kalmutzki, & Diercks, 2019). Since the discovery of the first MOF in 1999 (Li et al., 1999), research on MOF materials has been continuously increasing. Moreover, MOFs research has rapidly expanded into an array of real-life applications.

The widespread increase in MOF publications reveals a diversity of scientific approaches and research foci. As a result, researchers must manually assess the large amount of information related to this expanding field. Hence, there is a need for a concise approach to comprehend and engage in this field of study. This observation is particularly true for researchers new to the field, who must understand the historical and developing research landscape.

Scientometrics, the “quantitative study of science, communication in science, and



science policy" (Hess, 1997; Sengupta, 1992; Vinker, 1991), holds a distinct advantage for tackling researchers' needs. The ability to analyze large amounts of scholarly data, identify research trends and latent connections between disparate research topics (Sebastian, Siew, & Orimaye, 2017; Swanson, 1986) positions scientometric methods as a feasible approach. Due to its effectiveness, numerous scientometric analyses were previously conducted in economics and business studies (Bonilla, Merigo, & Torres-Abad, 2015; Castillo-Vergara, Alvarez-Marin, & Placencio-Hidalgo, 2018; Merigo & Yang, 2017; Ye et al., 2020), social studies (De Bakker, Groenewegen, & Den Hond, 2005; Guo et al., 2019), and especially biomedical studies (Guo et al., 2020; Krishnamoorthy, Ramakrishnan, & Devi, 2009). For example, Jeong et al. (2020) conducted a literature-based network analysis to examine the drugs and side effects across nearly 170,000 publications; Mak et al. (2022) explored the current landscape of AI-driven drug candidates for future development.

Scientometric analysis has been pursued in various material science areas, including MOF research. For example, Naseer et al. (2022) collected 1,187 publications and applied counting-based approaches in bibliometric analysis to explore research trends and theoretical approaches to wastewater decontamination. Ye and Yang (2022) collected 2,353 MOFs discussing the use of MOFs in electrochemistry. Shidiq (2023) analyzed 1,000 publications in the subarea of nano metal-organic frameworks in medical science to discover recurring phenomena and research themes. Another example is provided by Wang and Ho (2016), who explored the general landscape of MOFs by extracting terms from titles, author index and abstracts of collected articles and then manually performing analyses using a Microsoft Excel spreadsheet.

These previous investigations offer important insight into the MOF field, although they have several limitations: (1) The majority of these studies focus on a narrow subset of MOF research, so the complete landscape of MOF efforts remains unclear; (2) Several of the earlier analysis are based on small samples and representing work of a preliminary nature, and (3) Many of these analyses were conducted manually, relying solely on either human expertise or data perspectives.

Our investigation aims to provide a more in-depth analysis than previous research, to provide a comprehensive view of the MOF community. To meet this research aim, we developed a significantly larger corpus of publication data: publication data, including citation, article metadata and text, are drawn from 65,209 articles, which is more than ten-fold greater than previous studies. We also performed additional analysis to reveal yet-to-be-explored specific research areas and topics by applying network analysis and natural language processing techniques. Furthermore, a notable



contribution of this study is that we integrated viewpoints from both data and human experts. This approach addresses an important limitation in scientometric analysis (Lund, 2021). Many previous studies are either based solely on data or on human expertise, which can lead to a range of limitations. Without human expertise, valuable insights and links between different studies that are not apparent in data can be overlooked (Zanzotto, 2019), and the latter can be limited by relying on human expertise alone (especially when the domain is multidisciplinary). Hence, we designed our data analysis workflow with human-in-the-loop to deliver a comprehensive landscape of MOFs research. We first conducted data-driven exploratory analysis using scientometric methods, and then refined the data analysis by iterative engagement with domain expert scientists.

The four research objectives are to:

1. Identify the most influential research works and journals, and their latent connections they have to the area of metal-organic framework materials,
2. Determine major research directions in the MOF area, and their trends over time,
3. Reveal specific communities that exist in the research network, and
4. Recognize their detailed research topics.

In addition to the above research objectives, we explored the following question: Is the most cited study always the most impactful study? The sections that follow present our methods, findings, and discuss the research implications.

2 Method

To answer the research questions outlined above, we designed an expert knowledge guided scientometric analysis for discovering the landscape of the MOF community in three main steps: (1) data collection and processing, (2) bibliometric and network analysis, and (3) specific groups detection and topic modeling. Our domain collaborators, who are scientists in the field of metal-organic frameworks, were involved in each phase of the analysis as illustrated in Figure 1.

First, the target articles for data collection were validated by numerous scientists. Second, we confirmed that the scientometric research objective met the domain researchers' needs. Third, we pursued the initial scientometric data analysis. Finally, the initial results were assessed and refined by domain experts in two ways: the experts enhanced the initial results by adding insights and latent connections that are not apparent in the data, and provided suggestions to remove analysis components that are not helpful to domain researchers. For example, the list of organizations conducting different research topics was removed from the analysis, since researchers'



affiliations may change.

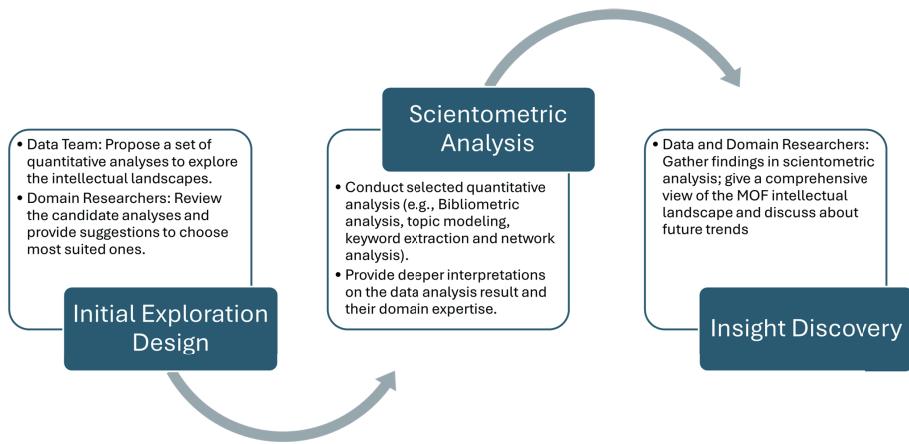


Figure 1. Overall workflow of the research design.

2.1 Data collection and processing

Our data collection and processing approach consists of two steps. First, we gather raw data related to MOF scientific publications. This data in raw form may consist of information regarding citation dependency, abstract and metadata from CrossRef (CrossRef, n.d.) and Scopus indexing systems (Scopus, n.d.). Secondly, we parse this raw data to a structured format. From this structured database, we then construct a research graph network for MOFs.

The data collection process was initiated by gathering MOF crystal structures included in the Cambridge Structural Database (CSD) (Groom et al., 2016), which is a database validated by materials scientists containing a wide range of crystal material structures. Our search strictly focused on the MOFCSD database (Moghadam et al., 2017), which is a subset of the CSD that contains “true MOFs.” Materials from related fields (e.g. 1-d coordination polymers) are not included. We used Digital Object Identifier (DOI) to identify publications associated with each collected MOF structure. Among 10,636 crystal structures in the database, 5,683 unique scientific articles along with their DOI were identified. For each identified publication, we also retrieved DOIs of papers in the reference list (in JSON - a structured text format) using CrossRef API. In total, we found 65,209 unique articles. After gathering a complete list of DOIs for articles above, we used these DOI to retrieve and download the metadata (e.g. title, journal, publication year, keywords, authors, affiliation) along with the abstract of each identified MOF article from Scopus in XML format. After parsing XML files collected from



Scopus, we found 112,115 unique authors, 9,018 unique affiliations and 60,652 unique indexed keywords closely related to the metal-organic framework studies. We linked the above metadata to their corresponding articles. The overall workflow of this data collection process is illustrated in Figure 2.

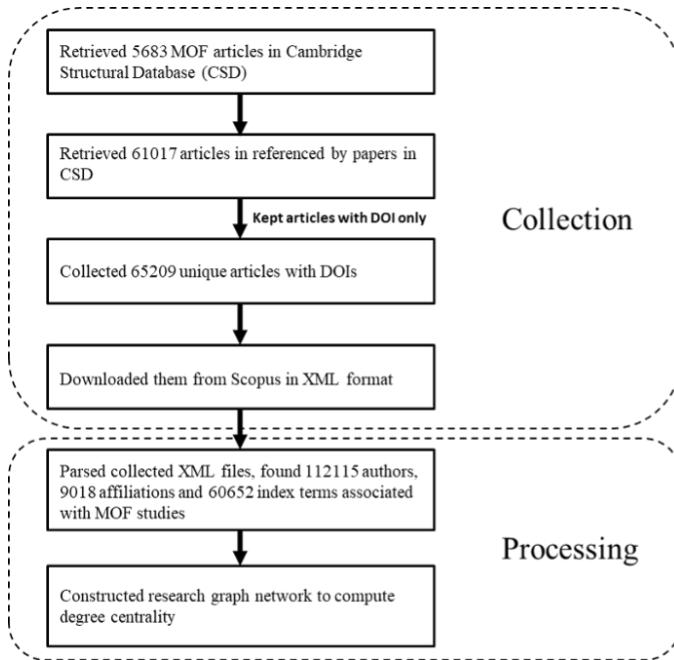


Figure 2. Overall workflow of data collection and processing.

2.2 MOF research network analysis

We analyzed MOF research networks from different angles: (1) impactful research entities, (2) change in topics over time, and (3) different research communities in the MOF area. The network analysis is performed by Python3 library *Networkx* (Hagberg, Swart, & SChult, 2008) and software tool *VOSviewer* (Van Eck & Waltman, 2010).

2.3 Most impactful entities in the field

In this study, impactful publications include works that are highly cited by other studies in the same research field. We constructed the MOF research network using *Networkx* as a directed graph, where each node represents one research publication, and edges represent citation dependency. To discover entities (e.g. articles, journals)



that produced impactful research output, we used degree centrality scores to quantify the article impact. In *Networkx*, the degree centrality of a node V is defined as the number of nodes it connects to (both in and out), then normalized by dividing the maximum possible degree of the network (Hagberg et al., 2008).

Based on this definition, a high degree centrality score for a publication indicates a significant connection to other research works. We computed the centrality score for every node in our MOF graph network and sorted scores in descending order. We also list the corresponding number of references of nodes for comparison.

2.4 Specific community detection

Communities, also called clusters in a graph network, consist of a set of nodes with similar features. The community detection (“clustering”) algorithm used in this paper is developed from a modularity-based algorithm (Clauset, Newman, & Moore, 2004) and is implemented by the software *VOSviewer* (Clauset et al., 2004; Van Eck & Waltman, 2010; Waltman, Van Eck, & Noyons, 2010).

To determine specific research groups in the field of MOFs, we conducted network co-citation analysis, which determines proximity of two publications by the number of times that a third publication cites them simultaneously.

2.5 Topic trends and modeling

As part of our analysis, we pursued text analysis, enhanced by domain experts, to detect clusters representing (1) research topic trends over time, and (2) detailed research topics in specific research directions. The first component is addressed using the occurrence of indexed terms and implemented by *VOSviewer*. Both unigram terms and n-gram terms are taken into account. The second component of the analysis is conducted using a noise-reducing topic modeling algorithm (Churchill & Singh, 2021) developed based on Latent Dirichlet Allocation (LDA). We gathered the associated abstracts of papers for each community detected, and then applied the topic modeling algorithm on the set of abstracts. Candidate topics generated by the algorithm are analyzed by both data and domain experts.

3 Results and findings

For preliminary data exploration, we plotted the publication distribution by year and the citation-rank distribution in Figure 3, to see the publication and citation pattern.

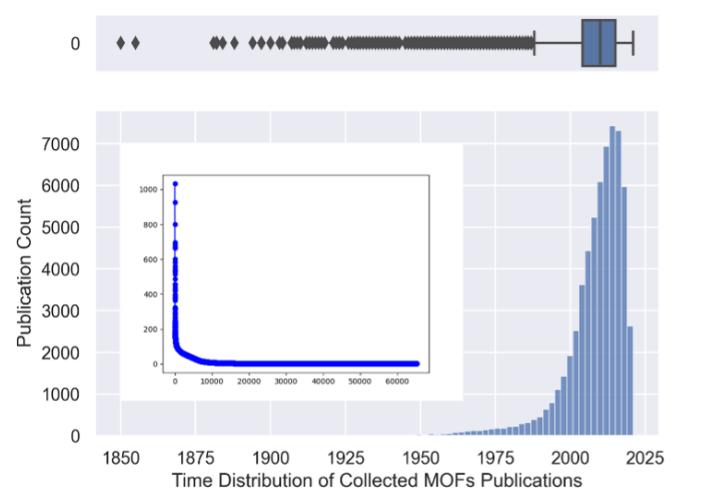


Figure 3. Publication data exploration in metal-organic framework area.

The publication distribution resembles a Zipf distribution function (Egghe, 2005; Perc, 2010) with the maximum number of publications peaking in 2013 (near 7,000 publications). From approximately 2000, the number of MOF publications began growing exponentially, with a significant number of research works being published – 75% of articles between 2005 to 2020. Pre-2000 publications refer primarily to “coordination polymers.” The term “MOF” emerged in 1999 and refers to those coordination polymers that retain a permanently open porous architecture even after guest/solvent removal (Li et al., 1999). Among these publications, their citation-rank distribution seems to be extremely skewed to the right (Figure 3, inset)—this skewness indicates that only a small percentage of publications receive most citations, whereas many publications are cited less frequently, consistent with a Pareto distribution. Next, we ask: what are the characteristics of these highly cited papers? Are these the most influential in the domain science?

3.1 Most impactful entities

To identify influential MOF research, we computed the centrality value of each node inside the research network (each node represents one scientific publication in the MOFs area); we sorted the graph nodes based on the centrality value in descending order. The top 20 most influential publications are listed in Table 1.



Table 1. Top 20 most influential publications based on degree centrality.

Publication Title	Journal Name	Year	Type	Referenced By	Centrality
Single-crystal structure validation with the program PLATON (Spek, 2003)	Journal of Applied Crystallography	2003	Article	16,311	0.0158
Functional porous coordination polymers (Kitagawa et al., 2004)	Angewandte Chemie - International Ed.	2004	Review	9,724	0.0142
A short history of SHELX (Sheldrick, 2008)	Acta Crystallographica Section A: Foundations of Crystallography	2008	Review	80,558	0.0123
Reticular synthesis and the design of new materials (Yaghi et al., 2003)	Nature	2003	Review	8,012	0.0107
Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage (Eddaoudi et al., 2002)	Science	2002	Article	6,717	0.0104
Luminescent functional metal-organic frameworks (Cui et al., 2012)	Chemical Reviews	2012	Review	4,960	0.0102
Selective gas adsorption and separation in metal-organic frameworks (Li et al., 2009)	Chemical Society Reviews	2009	Article	6,936	0.0102
Carbon dioxide capture in metal-organic frameworks (Sumida et al., 2012)	Chemical Reviews	2012	Article	5,090	0.0092
Metal-organic frameworks for separations (Li et al., 2012)	Chemical Reviews	2012	Review	5,284	0.0089
Luminescent metal-organic frameworks (Allendorf et al., 2009)	Chemical Society Reviews	2009	Article	4,407	0.0086
Metal-organic framework materials as catalysts (Lee et al., 2009)	Chemical Society Reviews	2009	Article	6,826	0.0083
Metal-organic framework materials as chemical sensors (Kreno et al., 2012)	Chemical Reviews	2012	Review	5,692	0.0082
Modular chemistry: Secondary building units as a basis for the design of highly porous and robust metal-organic carboxylate frameworks (Eddaoudi et al., 2001)	Accounts of Chemical Research	2001	Article	5,007	0.0081
From molecules to crystal engineering: Supramolecular isomerism and polymorphism in network solids (Moulton & Zaworotko, 2001)	Chemical Reviews	2001	Review	6,437	0.0079
Hybrid porous solids: Past, present, future (Ferey, 2008)	Chemical Society Reviews	2008	Article	5,139	0.0075
Interpenetrating Nets: Ordered, Periodic Entanglement (Batten & Robson, 1998)	Angew Chem Int Ed Engl	1998	Review	3,859	0.0070
Hydrogen storage in metal-organic frameworks (Murray et al., 2009)	Chemical Society Reviews	2009	Article	3,954	0.0070
The chemistry and applications of metal-organic frameworks (Furukawa et al., 2013)	Science	2013	Review	9,023	0.0067
A homochiral metal-organic porous material for enantioselective separation and catalysis (Seo et al., 2000)	Nature	2000	Article	3,730	0.0065
Design and synthesis of an exceptionally stable and highly porous metal- organic framework (Li et al., 1999)	Nature	1999	Article	6,486	0.0064



By analyzing the top 20 most influential publications above, we find that they break down into three categories: (1) technical software programs (crystallography), (2) original synthesis experiments, and (3) research reviews focusing on general applications and concepts. Technical software papers include those in which crystallography programs are cited. These papers are foundational in that crystal structure elucidation and validation are important components of any publication regarding MOFs. These include publications by Spek's work on single-crystal structure validation with the PLATON program (Spek, 2003), and Sheldrick's short history of SHELIX (Sheldrick, 2008). In our view, Speck is cited more than Sheldrick because it is a historical review on SHELX, crystallographic program to solve structures, while PLATON is more impactful because it is used to validate CIF files (Crystallographic Information Files, the data exchange standard file format used in materials research) (Hall, Allen, & Brown, 1991) before they can be accepted into the CSD (Cambridge Structural Database, as explained above).

Note that Table 1 shows the reference count is not always aligned with the centrality value—for example, the study conducted by Sheldrick is referenced many more times than the study by Spek, although the latter study has higher centrality.

Original synthesis foundations include the preparation and discovery of new MOFs. This data set includes the works by Li (1999), Eddaoudi (2002), and Seo (2000). The work by Li (1999) presented MOF-5 and introduces the importance of the polyoxometalate clusters in the formation of strong bonds, directionality, and the use of building blocks for forming targeted frameworks. This discovery led to the first demonstration of permanent micro-porosity in MOFs and is considered the beginning of the field. Eddaoudi's two papers (Eddaoudi et al., 2001; 2002) introduce the isoreticular principle in the form of isorecticular expansion and functionalization, which represents the first example of chemical control over the topology of a framework. Eddaoudi also emphasizes the importance of building blocks from a geometric perspective and the ability of the researcher to predict and thus design framework topologies. Seo's work presents a homochiral MOF for enantioselective separation and catalysis which embodies the fundamentals presented in the topological design of Eddaoudi and solid-state synthesis properties achieved by from the topology itself (Seo et al., 2000). Seo's paper is highly influential since not only does it provide the first structure property relationship in the field of MOFs, it also opens the door to catalysis, a highly studied and important chemical property. These four fundamental papers integrate the use of the SBUs (secondary building units, key components in MOF structure) and the isoreticular principle. This addresses anomalies found in coordination polymers and zeolites, including lack of permanent porosity in coordination polymers, lack of general functionalization in



Research Papers

zeolites as well as the inability to predict crystal structures. This paradigm shift highlights the central theorem in MOFs: the expansion of solid-state crystal engineering with encoded properties.

Review papers break down into either foundational (conceptual) or applications categories. The foundational review by Yaghi (2003) defines reticular chemistry as a logical approach “to the synthesis of robust materials with pre-designed building blocks, extended structures, and properties.” Here, Yaghi proposes that MOFs can be considered to be a subclass of crystal engineering. The review highlights the challenges from the pre-paradigm perspective by posing two questions.

1. First, of the almost unlimited possible combinations, which can be expected to form and how can they be synthesized?
2. Second, with few exceptions, MOFs based on M-N linkages in which the vertex of the network is simply a single atom tend to form structures which collapse upon removal of ‘guest’ atoms from the pores, rendering the structure nonporous. How then can we ensure that our structures will be robust?

The first question was answered by considering the geometries of the SBUs and linkers (connections between molecules, components in MOF structure) as well as the symmetry of the structure and the number of edges and vertices contained within the structure. This is a fundamentally important concept called the minimal transitivity principle and determines which default structures are formed. The second question is discussed by assessing rigidity and directionality of the polycarboxylate SBUs, their adopted geometries, as well as the ability of the reactants to retain structure under synthesis conditions. An important aspect of the review addressed pore size capacity in relation to Zeolites. Overall, Yaghi discusses complexity in terms of non-default structures and provides a conceptual approach to aid the designer in MOF synthesis.

The second most impactful conceptual review is written by Furukawa in 2013 (Furukawa et al., 2013). MOF research had advanced for over a decade by this time. The review also examines the progress of the synthesis of ultrahigh porosity of MOFs over time. Post-synthetic modification strategies, multivariate approaches, and topological designs which have aided in the selectivity, complexity, and synergistic effects in MOFs were highlighted. The resulting applications from this design space, such as gas storage, separation, catalysis, fuel cells, super capacitors, membrane, and thin film applications were highlighted. Thus, this marks the evolution of the field to that date. Other foundational reviews include those by Kitagawa (2004), Moulton (2001), Ferey (2008), and Batten (1998); Batten’s review (1998) covers coordination polymers; Kitagawa’s on MOF design (2004), which



included a scientometric analysis of a number of publications on coordination polymers by year from 1990 to 2002; Moulton's paper (2001) on polymorphism, distinguishing between crystal engineering and crystal structure prediction; and Ferey's work regarding SBUs in zeolites versus highly porous MOFs (Ferey, 2008). These reviews provide the foundational understanding required to access various nets in MOFs. The reviews examining applications are this by Cui (Cui et al., 2012) on luminescent functional metal organic frameworks, Li's on selective gas adsorption and separation in metal-organic frameworks (Li, Kuppler, & Zhou, 2009), Lee's (2009) on metal organic framework materials as catalysts, Sumida's (2012) on carbon dioxide capture in metal organic frameworks, Li's (2012) on metal-organic frameworks for separations, Allendorf's (2009) on luminescent metal organic frameworks, and Kreno's (2012) on metal organic framework material as chemical sensors, and hydrogen storage in metal organic frameworks.



Figure 4. Journals publishing highly impactful research.



The journals where impactful MOF studies are published are listed in Figure 4. We find that most impactful journals in MOFs are domain-specific (e.g. CrystEngComm, Crystal Growth and Design), instead of high-impact journals in general (e.g. Science, Nature).

3.2 Topic trend over time

We extracted research key terms from abstracts, and then constructed a keyword network. Figure 5 is illustrates the keyword network - each node represents one term, and its size represents its occurrence. Colored nodes represent articles in the same community.

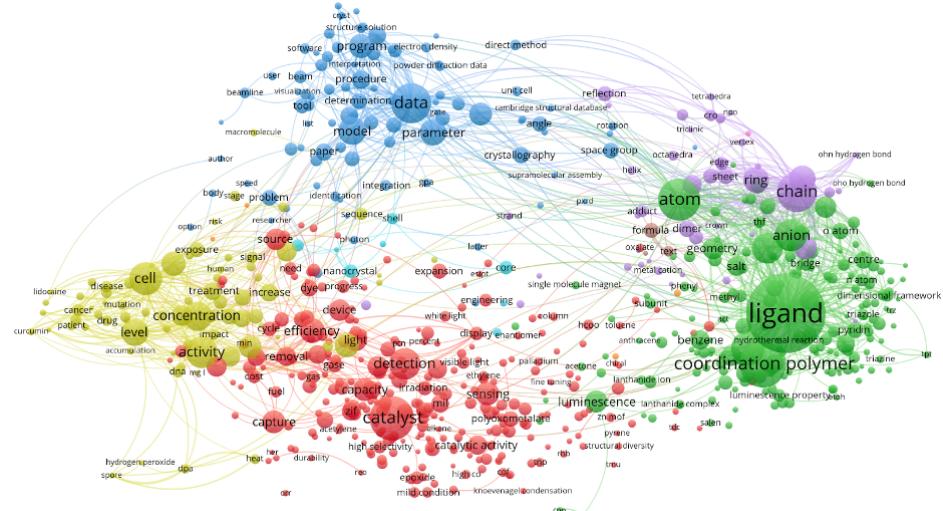


Figure 5. Keywords extracted from MOF research articles and potential communities.

Based on the keyword network above, four main communities are detected. By analyzing terms within each community, we determined that these four communities point to four main MOF research topics: (1) MOF synthesis, (2) properties/applications of MOF materials, (3) use of MOF in biomedicine, and (4) MOF data processing and modeling. A list of example terms in each community is provided in Table 2.

Table 2. Example terms from the above four main research communities.

Properties	Biomedicine	Data Processing & Modeling	Synthesis Experiment
Catalyst	Cell	Data	Ligand
Uptake capacity	Cytotoxicity	Program	Hydrothermal condition
Chemical stability	Inhibitor	Processing	Single Crystal X-ray diffraction
Permanent porosity	Mutation	Software	Elemental analysis
Durability	Protein	Interface	Topology
Low cost	Treatment	Model	Solvothermal reaction

We examined the publication year of the key research terms to help understand

research trends in the MOF field. In Figure 6, each key term in the network is colored by its average publication year. The result suggests that MOF data modeling and using MOF materials in biomedicine have attracted researchers' attention since early years; more recently, research trends seem to have moved to discovery of effective synthesis strategies of MOF materials, as well as the properties and potential application of synthesized MOFs around the year of 2014.

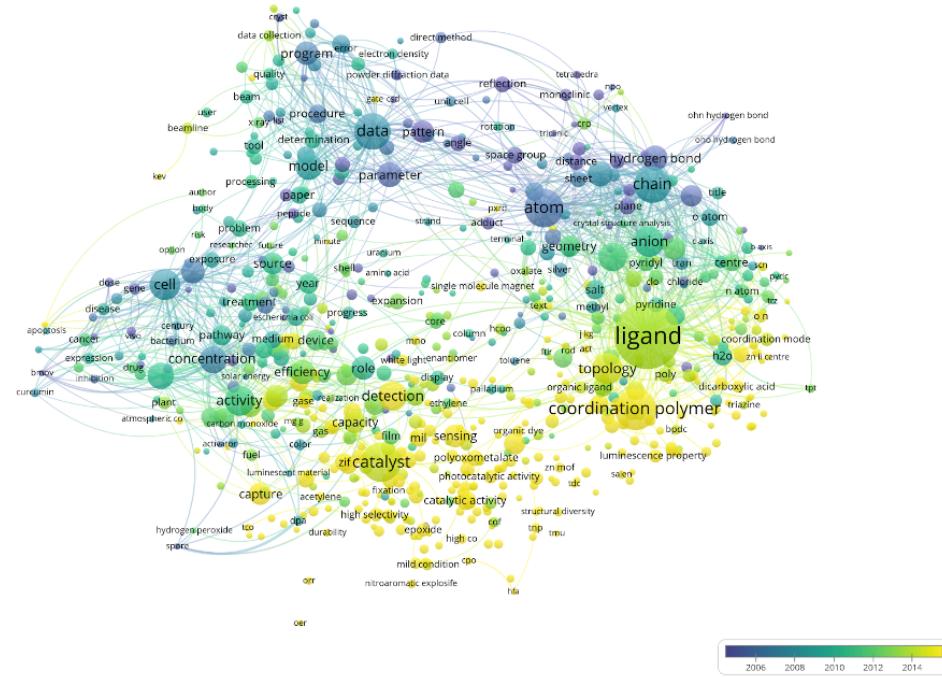


Figure 6. Trend of research topics over time.

3.3 Document co-citation analysis and topic modeling for research communities

We used co-citation analysis to further explore specific groups in the MOF area, and then applied topic modeling techniques and indexed terms to reveal research interests of each specific group. To construct a co-citation network, we limited the minimum number of citations of a cited reference to five, keeping the largest connected component. The resulting network is reported in Figure 7. In a manner similar to the high-level topic trend analysis reported above, we detected eight distinct specific groups in the network.



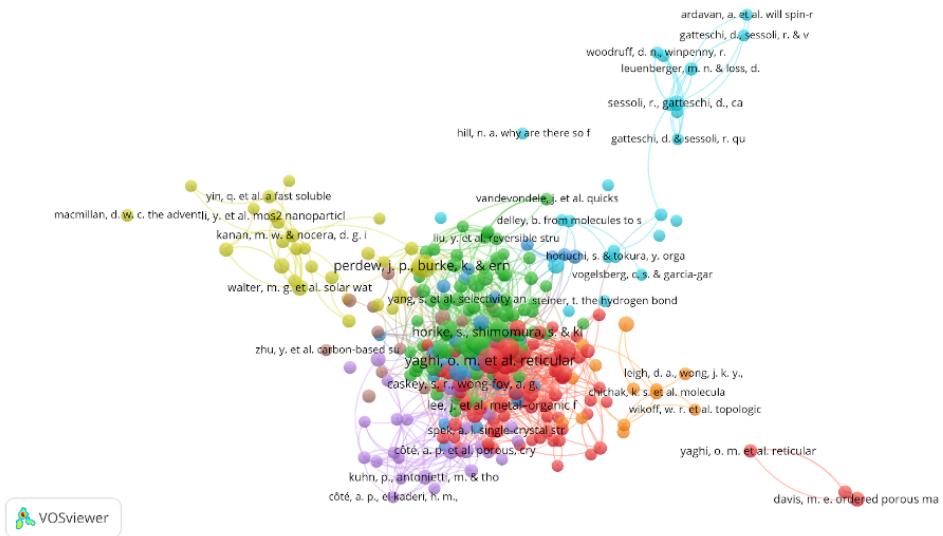


Figure 7. Potential specific research groups recognized by network analysis.

To examine specific research topics more closely, we applied noise-reducing LDA-based topic modeling algorithms (Churchill & Singh, 2021) to identify topics for each specific group. Indexed terms in each article are also considered. We also identified topics from each specific group. The results show at least eight specific MOF research directions: (1) drug delivery, (2) zeolitic imidazolate frameworks, (3) Carbon capture and storage, (4) Catalytic materials, (5) homochiral MOFs and asymmetric catalysis, (6) magnetic MOF materials, (7) self-assembly MOFs, and (8) study of the electrical properties of MOFs.

4 Discussion

As an emerging material, MOFs are garnering significant interest from the scientific community, driven by their exceptional potential and unique material characteristics. Scientometric methods play a crucial role to not only better understand the current knowledge structure, but also advance the study of metal-organic frameworks. MOF research, at the intersection of chemistry, materials science, medicine and engineering, produces a vast body of literature in different scientific domains. Scientometric analysis is uniquely positioned to handle this complexity, by using data-driven quantitative methods to dissect and examine the extensive volume of research.

While general methods supporting scientometric analysis, such as network

analysis and NLP, can reveal important information, an exclusively data-driven approach lacks human expert feedback. Research shows that engaging domain experts provides greater insight and enriches data analysis (Wu et al., 2022; Xin et al., 2018; Zanzotto, 2019). A significant contribution of the research presented in this paper is the iterative engagement of the domain experts, specifically research scientists, who helped reveal a more comprehensive view of the MOF research intellectual landscape. Human experts bring domain knowledge to the study allowing greater depth of analysis. In our analysis we integrated their knowledge about historical developments, current bottlenecks and future implications. In addition, given the complexity of scientific studies such as those on MOFs, we found that some connections between disparate research efforts may be overlooked by purely data-based analysis. Similarly, subtle aspects are often not seen when they are not immediately apparent in raw data. Domain experts' involvement allowed us to mitigate the above limitations and supported a more in-depth interpretation on top of data analysis result by evaluating the quality and significance of specific pieces of studies. They also provided insights on how different studies, even from various disciplines, could enrich each other. Finally, human scientists can suggest innovative methodologies as well as future research direction based on the initial data analysis result, including those that are not obvious from data alone.

5 Conclusions

In this study, we demonstrate the value of scientometric methods empowered by human expertise as a means for understanding the MOF field and its evolution. We incorporated both data and human aspects while discovering a more comprehensive view of MOFs research. We first conducted data-driven scientometric analysis, including bibliometric, network and text analysis, and then expanded the initial results by applying domain expertise from human researchers to develop a detailed, comprehensive analysis of intellectual landscape of metal-organic frameworks studies.

Specifically, by integrating the result of scientometric analysis and expertise from domain scientists, (1) we identified most the impactful journals and research works along with their interconnections in the MOF area, (2) we determined the major research directions and their trends over time using a clustering algorithm, and (3) we discovered specific research groups and topics in the MOF research network.

Compared to previous studies, our work greatly improves the intellectual landscape of MOFs study in the following ways: (1) According to our review of available information, the number of publications included in our analysis exceeds the volume gathered in other studies more than tenfold, thus improving understanding of the



MOF research landscape, (2) We applied network analysis and natural language processing techniques to obtain data-driven exploration results, and (3) we further enhanced the data exploration result by human expertise to identify nuances overlooked by data alone.

In the future, we plan to further examine research topics revealed by this study. For example, we are interested in extracting information from the scientific literature to understand potential connections between synthesis routes and resulting material characteristics.

Funding information

This work is funded by NSF OAC # 2118201.

Acknowledgements

We acknowledge and are grateful for the support from MOF domain scientists from the Reticular Synthesis and Materials Design Lab, University of Central Florida, as well as the Computational Design of Materials for Energy Applications (CoDe MatE) Lab in Colorado School of Mines.

Author contributions

Xintong Zhao (xz485@drexel.edu): Data curation (Equal), Formal analysis (Lead), Investigation (Equal), Writing – original draft (Lead);

Kyle Langois (kylerlanglois@knights.ucf.edu): Formal analysis (Equal), Investigation (Equal), Methodology (Equal), Validation (Equal);

Jacob Furst (jfurst@knights.ucf.edu): Formal analysis (Equal), Investigation (Equal), Methodology (Equal), Validation (Equal);

Yuan An (ya45@drexel.edu): Conceptualization (Equal), Methodology (Equal), Supervision (Equal);

Xiaohua Hu (xh29@drexel.edu): Conceptualization (Equal), Methodology (Equal), Supervision (Equal);

Diego Gomez Gualdrón (dgomezgualdrón@mines.edu): Conceptualization (Equal), Formal analysis (Supporting), Funding acquisition (Equal), Investigation (Equal), Supervision (Equal);

Fernando Uribe-Romo (fernando@ucf.edu): Conceptualization (Equal), Funding acquisition (Equal), Investigation (Equal), Methodology (Equal), Supervision (Equal), Writing – review & editing (Equal);

Jane Greenberg (jg3243@drexel.edu): Conceptualization (Equal), Funding acquisition (Equal), Investigation (Equal), Supervision (Equal), Writing – review &



References

Allendorf, M. D., Bauer, C. A., Bhakta, R., & Houk, R. (2009). Luminescent metal-organic frameworks. *Chemical Society Reviews*, 38(5), 1330–1352.

Batten, S. R., & Robson, R. (1998). Interpenetrating nets: ordered, periodic entanglement. *Angewandte Chemie International Edition*, 37(11), 1460–1494.

Bonilla, C. A., Merig'o, J. M., & Torres-Abad, C. (2015). Economics in latin america: a bibliometric analysis. *Scientometrics*, 105, 1239–1252.

Castillo-Vergara, M., Alvarez-Marin, A., & Placencio-Hidalgo, D. (2018). A bibliometric analysis of creativity in the field of business economics. *Journal of Business Research*, 85, 1–9.

Cavka, J. H., Jakobsen, S., Olsbye, U., Guillou, N., Lamberti, C., Bordiga, S., & Lillerud, K. P. (2008). A new zirconium inorganic building brick forming metal organic frameworks with exceptional stability. *Journal of the American Chemical Society*, 130(42), 13850–13851.

Churchill, R., & Singh, L. (2021). Topic-noise models: Modeling topic and noise distributions in social media post collections. In 2021 IEEE International Conference on Data Mining (ICDM) (pp. 71–80).

Clauset, A., Newman, M. E., & Moore, C. (2004). Finding community structure in very large networks. *Physical Review E*, 70(6), 066111.

Cote, A. P., Benin, A. I., Ockwig, N. W., O'Keeffe, M., Matzger, A. J., & Yaghi, O. M. (2005). Porous, crystalline, covalent organic frameworks. *Science*, 310(5751), 1166–1170.

Crossref.org. (n.d.). Retrieved October 31, 2023 from <https://www.crossref.org/>

Cui, Y., Yue, Y., Qian, G., & Chen, B. (2012). Luminescent functional metal- organic frameworks. *Chemical Reviews*, 112(2), 1126–1162.

De Bakker, F. G., Groenewegen, P., & Den Hond, F. (2005). A bibliometric analysis of 30 years of research and theory on corporate social responsibility and corporate social performance. *Business & Society*, 44(3), 283–317.

Dong, B., Xu, G., Luo, X., Cai, Y., & Gao, W. (2012). A bibliometric analysis of solar power research from 1991 to 2010. *Scientometrics*, 93(3), 1101–1117.

Eddaoudi, M., Kim, J., Rosi, N., Vodak, D., Wachter, J., O'Keeffe, M., & Yaghi, O. M. (2002). Systematic design of pore size and functionality in isoreticular mofs and their application in methane storage. *Science*, 295(5554), 469–472.

Eddaoudi, M., Moler, D. B., Li, H., Chen, B., Reineke, T. M., O'keeffe, M., & Yaghi, O. M. (2001). Modular chemistry: secondary building units as a basis for the design of highly porous and robust metal- organic carboxylate frameworks. *Accounts of Chemical Research*, 34(4), 319–330.

Egghe, L. (2005). The power of power laws and an interpretation of lotkaian informetric systems as self-similar fractals. *Journal of the American Society for Information Science and Technology*, 56(7), 669–675.

Ferey, G. (2008). Hybrid porous solids: past, present, future. *Chemical Society Reviews*, 37(1), 191–214.



Furukawa, H., Cordova, K. E., O'Keeffe, M., & Yaghi, O. M. (2013). The chemistry and applications of metal-organic frameworks. *Science*, 341(6149), 1230444.

Ginsberg, A. (1990). Early transition metal polyoxoanions. *Inorganic Syntheses*, 71, 27.

Groom, C. R., Bruno, I. J., Lightfoot, M. P., & Ward, S. C. (2016, April). *The cambridge structural database* (Vol. 72) (No. 2). International Union of Crystallography (IUCr). Retrieved from <http://dx.doi.org/10.1107/S2052520616003954>

Guo, Y., Hao, Z., Zhao, S., Gong, J., & Yang, F. (2020). Artificial intelligence in health care: bibliometric analysis. *Journal of Medical Internet Research*, 22(7), e18228.

Guo, Y.-M., Huang, Z.-L., Guo, J., Li, H., Guo, X.-R., & Nkeli, M. J. (2019). Bibliometric analysis on smart cities research. *Sustainability*, 11(13), 3606.

Hagberg, A., Swart, P., & S Chult, D. (2008). *Exploring network structure, dynamics, and function using networkx* (Tech. Rep.). Los Alamos National Lab. (LANL), Los Alamos, NM (United States).

Hall, S. R., Allen, F. H., & Brown, I. D. (1991). The crystallographic information file (cif): a new standard archive file for crystallography. *Acta Crystallographica Section A: Foundations of Crystallography*, 47(6), 655–685.

Hess, D. J. (1997). *Science studies: An advanced introduction*. NYU Press.

Ho, Y.-S. (2014). A bibliometric analysis of highly cited articles in materials science. *Current Science*, 1565–1572.

Jeong, Y. K., Xie, Q., Yan, E., & Song, M. (2020, February). *Examining drug and side effect relation using author-entity pair bipartite networks* (Vol. 14) (No. 1). Elsevier BV. Retrieved from <http://dx.doi.org/10.1016/j.joi.2019.100999>

Kinoshita, Y., Matsubara, I., & Saito, Y. (1959). The crystal structure of bis (succinonitrilo) copper (i) nitrate. *Bulletin of the Chemical Society of Japan*, 32(7), 741–747.

Kitagawa, S., Kitaura, R., & Noro, S.-i. (2004). Functional porous coordination polymers. *Angewandte Chemie International Edition*, 43(18), 2334–2375.

Kreno, L. E., Leong, K., Farha, O. K., Allendorf, M., Van Duyne, R. P., & Hupp, J. T. (2012). Metal–organic framework materials as chemical sensors. *Chemical Reviews*, 112(2), 1105–1125.

Krishnamoorthy, G., Ramakrishnan, J., & Devi, S. (2009). Bibliometric analysis of literature on diabetes (1995–2004).

Lan, J., Wei, R., Huang, S., Li, D., Zhao, C., Yin, L., & Wang, J. (2022). Indepth bibliometric analysis on research trends in fault diagnosis of lithiumion batteries. *Journal of Energy Storage*, 54, 105275.

Lee, J., Farha, O. K., Roberts, J., Scheidt, K. A., Nguyen, S. T., & Hupp, J. T. (2009). Metal–organic framework materials as catalysts. *Chemical Society Reviews*, 38(5), 1450–1459.

Li, H., Eddaoudi, M., O'Keeffe, M., & Yaghi, O. M. (1999). Design and synthesis of an exceptionally stable and highly porous metal-organic framework. *Nature*, 402(6759), 276–279.

Li, J.-R., Kuppler, R. J., & Zhou, H.-C. (2009). Selective gas adsorption and separation in metal–organic frameworks. *Chemical Society Reviews*, 38(5), 1477–1504.

Li, J.-R., Sculley, J., & Zhou, H.-C. (2012). Metal–organic frameworks for separations. *Chemical Reviews*, 112(2), 869–932.



Lund, B. D. (2021). How to select better topics and design better bibliometrics and scientometrics studies: A perspective. *Journal of Scientometric Research*, 10(3), 348–351.

Mak, K.-K., Balijepalli, M. K., & Pichika, M. R. (2022). Success stories of ai in drug discovery—where do things stand? *Expert Opinion on Drug Discovery*, 17(1), 79–92.

Merigo, J. M., & Yang, J.-B. (2017). A bibliometric analysis of operations research and management science. *Omega*, 73, 37–48.

Moghadam, P. Z., Li, A., Wiggins, S. B., Tao, A., Maloney, A. G. P., Wood, P. A., ... Fairen-Jimenez, D. (2017, March). *Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future* (Vol. 29) (No. 7). American Chemical Society (ACS). Retrieved from <http://dx.doi.org/10.1021/acs.chemmater.7b00441>

Moulton, B., & Zaworotko, M. J. (2001). From molecules to crystal engineering: supramolecular isomerism and polymorphism in network solids. *Chemical Reviews*, 101(6), 1629–1658.

Murray, L. J., Dinc̄a, M., & Long, J. R. (2009). Hydrogen storage in metal– organic frameworks. *Chemical Society Reviews*, 38(5), 1294–1314.

Naseer, M. N., Jaafar, J., Junoh, H., Zaidi, A. A., Kumar, M., Alqahtany, A., ... & Aldossary, N. A. (2022). Metal-organic frameworks for wastewater decontamination: Discovering intellectual structure and research trends. *Materials*, 15(14), 5053.

Perc, M. (2010). Zipf's law and log-normal distributions in measures of scientific output across fields and institutions: 40 years of slovenia's research as an example. *Journal of Informetrics*, 4(3), 358–364.

Scopus. (n.d.). Retrieved from <https://www.scopus.com/home.uri>

Sebastian, Y., Siew, E.-G., & Orimaye, S. O. (2017). *Emerging Approaches in Literature-based Discovery: Techniques and Performance Review* (Vol. 32). Cambridge University Press (CUP). Retrieved from <http://dx.doi.org/10.1017/S0269888917000042>

Sengupta, I. N. (1992). Bibliometrics, informetrics, scientometrics and librametrics: an overview.

Seo, J. S., Whang, D., Lee, H., Jun, S. I., Oh, J., Jeon, Y. J., & Kim, K. (2000). A homochiral metal– organic porous material for enantioselective separation and catalysis. *Nature*, 404(6781), 982–986.

Sheldrick, G. M. (2008). A short history of shelx. *Acta Crystallographica Section A: Foundations of Crystallography*, 64(1), 112–122.

Shidiq, A. P. (2023). A bibliometric analysis of nano metal-organic frameworks synthesis research in medical science using vosviewer. *ASEAN Journal of Science and Engineering*, 3(1), 31–38.

Spek, A. (2003). Single-crystal structure validation with the program platon. *Journal of Applied Crystallography*, 36(1), 7–13.

Sumida, K., Rogow, D. L., Mason, J. A., McDonald, T. M., Bloch, E. D., Herm, Z. R., ... Long, J. R. (2012). Carbon dioxide capture in metal– organic frameworks. *Chemical Reviews*, 112(2), 724–781.

Swanson, D. R. (1986). Undiscovered public knowledge. *The Library Quarterly: Information, Community, Policy*, 56(2), 103–118. Retrieved 2023-12-11, from <http://www.jstor.org/stable/4307965>

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



Research Papers

Vinker, P. (1991). Possible main criteria of the impact of publications of science. *Dr. N. Sengupta's International Festschrift Volume (communicated)*.

Waltman, L., Van Eck, N. J., & Noyons, E. C. (2010). A unified approach to mapping and clustering of bibliometric networks. *Journal of Informetrics*, 4(4), 629–635.

Wang, C., Liu, X., Demir, N. K., Chen, J. P., & Li, K. (2016). Applications of water stable metal–organic frameworks. *Chemical Society Reviews*, 45(18), 5107–5134.

Wang, C.-C., & Ho, Y.-S. (2016). Research trend of metal–organic frameworks: a bibliometric analysis. *Scientometrics*, 109, 481–513.

Wu, X., Xiao, L., Sun, Y., Zhang, J., Ma, T., & He, L. (2022). A survey of human-in-the-loop for machine learning. *Future Generation Computer Systems*, 135, 364–381.

Xin, D., Ma, L., Liu, J., Macke, S., Song, S., & Parameswaran, A. (2018). Accelerating human-in-the-loop machine learning: Challenges and opportunities. In *Proceedings of the Second Workshop on Data Management for End-to-End Machine Learning* (pp. 1–4).

Yaghi, O. M., Kalmutzki, M. J., & Diercks, C. S. (2019). *Introduction to Reticular Chemistry: Metal-Organic Frameworks and Covalent Organic Frameworks*. John Wiley & Sons.

Yaghi, O. M., O'Keeffe, M., Ockwig, N. W., Chae, H. K., Eddaoudi, M., & Kim, J. (2003). Reticular synthesis and the design of new materials. *Nature*, 423(6941), 705–714.

Ye, N., Kueh, T.-B., Hou, L., Liu, Y., & Yu, H. (2020). A bibliometric analysis of corporate social responsibility in sustainable development. *Journal of Cleaner Production*, 272, 122679.

Ye, W., & Yang, W. (2022). Exploring metal-organic frameworks in electrochemistry by a bibliometric analysis. *Journal of Industrial and Engineering Chemistry*, 109, 68–78.

Zanzotto, F. M. (2019). Human-in-the-loop artificial intelligence. *Journal of Artificial Intelligence Research*, 64, 243–252.



Copyright: © 2024 Xintong Zhao, Kyle Langlois, Jacob Furst, Yuan An, Xiaohua Hu, Diego Gomez Gualdrón, Fernando Uribe-Romo, Jane Greenberg. Published under a Creative Commons Attribution 4.0 International (CC BY 4.0) license.

