# A Survey of Decomposition-Based Evolutionary Multi-Objective Optimization: Part II—A Data Science Perspective

Mingyu Huang[1] and Ke Li[2]

[1]School of Computer Science and Engineering, UESTC, Chengdu, PR China
[2]Department of Computer Science, University of Exeter, EX4 4QF, Exeter, UK
m.huang.gla@outlook.com; k.li@exeter.ac.uk



**Abstract:** This paper presents the second part of the two-part survey series on decomposition-based evolutionary multi-objective optimization where we mainly focus on discussing the literature related to multi-objective evolutionary algorithms based on decomposition (MOEA/D). Complementary to the first part, here we employ a series of advanced data mining approaches to provide a comprehensive anatomy of the enormous landscape of MOEA/D research, which is far beyond the capacity of classic manual literature review protocol. In doing so, we construct a heterogeneous knowledge graph that encapsulates more than 5,400 papers, 10,000 authors, 400 venues, and 1,600 institutions for MOEA/D research. We start our analysis with basic descriptive statistics. Then we delve into prominent research/application topics pertaining to MOEA/D with state-of-the-art topic modeling techniques and interrogate their sptial-temporal and bilateral relationships. We also explored the collaboration and citation networks of MOEA/D, uncovering hidden patterns in the growth of literature as well as collaboration between researchers. Our data mining results here, combined with the expert review in Part I[1], together offer a holistic view of the MOEA/D research, and demonstrate the potential of an exciting new paradigm for conducting scientific surveys from a data science perspective.
**Keywords:** Multi-objective optimization, decomposition, data mining, topic modeling, network analysis, data visualization.

## 1 Introduction

In the first part of this survey series, we have developed a major methodological landscape of decomposition-based evolutionary multi-objective optimization (EMO), particularly multi-objective evolutionary algorithm based on decomposition (MOEA/D) [1], from the perspective of a domain expert with more than 15 years' experience. While our survey in Part I adheres to a systematic literature review (SLR) protocol developed in [2], this approach, despite its thoroughness in terms of a methodological rigor, has yet to capture the comprehensive landscape across thousands of relevant articles. Several aspects remain unexplored, for instance:

- *What types of interdisciplinary research have been facilitated by MOEA/D?*
- *What are the emerging trends pertaining to MOEA/D research and its applications?*
- *How are the research agendas in MOEA/D linked with each other or applications?*

Moreover, for audiences who are relatively new to this field, they might also be interested in the following question:

- *What are the most active researchers or most prestigious venues in the MOEA/D community?*

[1]The article of Part I can be downloaded from this link, while the Appendix file can be found in this link. Code is available at this link

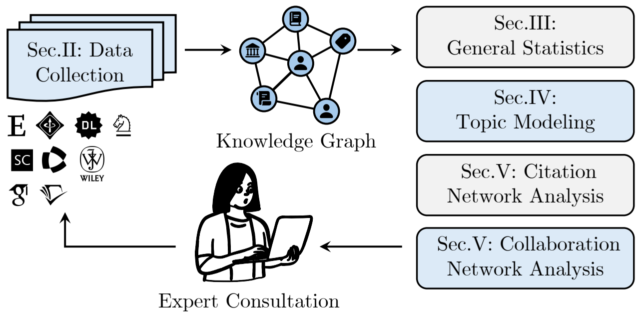


*Figure 1.* High-level schematic workflow of our data-driven literature survey.

There are many more such inquiries whose exploration can be very important for charting future research trajectories. For instance, understanding major applications can be a springboard for catalyzing new research challenges, while insights into the methodological evolution can equip researchers and practitioners from various disciplines with the knowledge pertinent to state-of-the-art algorithmic advancements. Nonetheless, given the daunting volume of literature within the niche of decomposition-based EMO, comprehensively synthesizing this body of work exceeds the capacity of any individual human expert and is susceptible to subjective biases [3].

In the past decade and beyond, we have seen an influx of exploratory analysis approaches for different forms of data, which significantly facilitate research in a wide range of disciplines. For example, topic modeling [4] from the natural language processing (NLP) community has assisted in extracting latent topics from large-scale text corpora like tweets on climate change [5] or hotel reviews [6]. Community detection [7] developed for complex networks are able to identify groups of individuals in social networks [8] or functional regions in brains [9]. More recently, the proliferation of large language models (LLMs) [10] has demonstrated unparalleled capability in performing labor-intensive tasks like data annotation [11] and wrangling [12] as well as text summarization [13]. Meanwhile, scholarly databases like Web of Science (WoS), Scopus and Semantic Scholar have incorporated tremendous amount of bibliographic entries with enriched metadata such as abstracts, publication venue and year, author information, and citation data, to name a few. This deluge of digital data on scholarly, combined with the abundance of data mining approaches, provides unprecedented opportunities not only for conducting large-scale mapping of the research landscape pertaining to specific domains [14–17], but also to uncover important patterns characterizing the structure and evolution of science [18–22].

**Contributions:** Bearing the above consideration in mind, in PART II, we developed an end-to-end workflow for conducting a data-driven literature review, a bold step forward beyond conventional SLR approaches. This workflow, sketched in Fig. 1, leverages a broad collection of data mining tools—from descriptive statistics and data visualization to advanced NLP techniques, network science approaches, and machine learning (ML) algorithms. We merged, processed, and structured diverse bibliographic data for $5,400+$ papers, $10,000+$ authors, $1,400+$ venues and $1,600+$ institutions in the decomposition-based EMO community as a heterogeneous knowledge graph, and applied this workflow to obtain an atlas of the research landscape by a series of dedicated retrospective analysis. Specifically:

- Section 2 introduces our **data collection** protocol, including the data sources, selected papers and the metadata relevant to our data analysis. Further, we also delineate the methods for data cleaning and the construction of the bibliographic knowledge graph. All these constitute the foundation for the subsequent analyses.
- By using the constructed knowledge graph, Section 3 conducts a **general analysis** of MOEA/D research on the trends and distributions of publications, venues, authors, and research disciplines.
- In Section 4, we systematically explore the underlying research topics and applications in MOEA/D literature by analyzing paper contents using **topic modeling**. We also interrogate the sptial-temporal distribution of these topics and interrogate their relationships with each other.

- In Section 5, we offer a complementary perspective for discovering research topics by exploring communities in the **citation network** of MOEA/D research. Further, we explore the evolutionary dynamics of this network as well as the disruptiveness of research. Additionally, we employ a main path analysis to identify landmark studies.
- In Section 6, we investigate several intriguing properties of the **collaboration network** of MOEA/D researchers. We also identify the most active researchers in the community, and explore the collaboration pattern among them.

Our data-driven workflow is further aided by EMO experts, wherein the results are iteratively validated and curated to align with their expertise via a human-AI collaboration paradigm.

**Related works:** While such practices have also been seen in some survey papers on other fields (e.g., [14–17]), they are either at 'narrow' breadth (e.g., focusing on a single type of data analysis) or 'shallow' depth (e.g., employing simple or out-dated methods). For instance, [14] and [17] restricted their reviews to citation network analysis, while [16] relied solely on topic modeling. [15] adopted both types of analysis, but they only reported basic network metrics and used non-negative matrix factorization for the topic model, which has been significantly outperformed by recent deep learning-based approaches [4]. In contrast, our literature mining workflow incorporates four analytical perspectives (see Fig. 1), each supported by a collection of state-of-the-art techniques.

In a another line, data-driven discovery in scholarly literature, often discussed under terms like 'science of science', 'scientometrics', or 'bibliometrics' [18–22], explores the characteristics of scientific research itself. We note that these works are orthogonal to our aim, as they primarily focus on general patterns, such as the sociology, psychology, economics, and information flow associated with science. Our work, in contrast, delves into the specific research landscape of decomposition-based EMO, MOEA/D in particular, offering a targeted exploration of this field. Beyond this, we also aim to embark on a paradigm shift for carrying out scientific surveys as a data science.

# 2 Data Collection

This section describes our data collection method and the construction of MOEA/D knowledge graph.

## 2.1 Candidate Study Identification

### 2.1.1 Data sources

Here we used the WoS for a systematic literature search and data extraction. WoS's extensive database encompasses $>$ 91M entries from 22,171 sources, including journals and conferences, and $>$ 500 publishers. This range covers well-known publishers such as IEEE, ACM, Elsevier, and Springer, as well as smaller but renown ones in computer science like PMLR and OpenReview. Choosing WoS is motivated by its diverse coverage and the quality of its metadata, a factor that has led to its frequent selection as the data foundation in many SLRs [23–25]. Note that unlike other works that directly scrape data from the web interface of WoS, we employed the `WoS API Expanded`. This approach ensures better reproducibility and flexibility, particularly critical for the large-scale data acquisition required by our study.

### 2.1.2 Search strategy

While traditional SLRs often use a combination of keywords and Boolean operators to identify relevant studies, our work aims to construct a comprehensive overview of the MOEA/D research landscape. Therefore, we initially included all papers that cited Zhang and Li's seminal MOEA/D paper [1]. Our rationale is that papers citing [1] are likely to either apply or extend MOEA/D, compare it with other EMO methods, or discuss it within a broader EMO context. We deemed all such papers, including those that only mention [1], as valuable for our analysis to accurately position MOEA/D within the entire EMO community. This initial filtering yielded a total of 5,606 works indexed by WoS as of March 2024, forming our set of candidate papers for further analysis.

**Remark 1**

The citation count for MOEA/D reported in WoS is lower than that found in other databases, such as Semantic Scholar (6,729) or Google Scholar (8,636). This discrepancy is primarily due to the broader inclusivity of the latter databases, which index a wider array of publication types, including arXiv preprints, patents, theses, and technical reports. While these sources increase the citation count, they may be less reliable nor relevant for our analysis, potentially introducing noise into the dataset. Therefore, they were excluded from consideration in this study.

## 2.2 Paper filtering

To ensure a comprehensive and unbiased analysis, our goal is to retain as many of the obtained candidate papers as possible, excluding them only for compelling reasons. To this end, we established the following five exclusion criteria:

- ✘ The paper is not written in English.
- ✘ Duplicated papers.
- ✘ The length of the paper is less than four pages (e.g., short or work-in-progress papers like GECCO Companion).
- ✘ Books, keynote records, non-published, workshop papers, and non-peer-reviewed manuscripts.
- ✘ Extended journal version of a conference paper.

**Remark 2**

We refrained from applying any exclusion criteria based on citation counts, as there is no universally accepted threshold for defining an 'influential paper'. Implementing such a criterion could inadvertently introduce publication bias into our results.

In summary, after applying these criteria, the final selection of papers was determined, comprising 5,404 papers published between 2007 and 2024.

## 2.3 Metadata Extraction and Wraggling

At the heart of our data-driven analysis lies a corpus of high-quality meta-features for each paper, which can be leveraged by data mining techniques to reveal latent patterns. We compiled a set of 13 meta-features from diverse sources, deemed essential for our analysis, as listed in Table 1.

### 2.3.1 General purpose features

WoS itself offers a wealth of metadata for each paper (refer to $F_1$-$F_{11}$), and we empirically found that its data quality significantly surpasses that of other databases like Semantic Scholar. For instance, Semantic Scholar exhibits a high rate (nearly 40%) of missing data in critical fields such as author affiliations and paper abstracts and lacks detailed information like keywords or subject categories. In contrast, only 113 out of the 5,404 papers in our dataset from WoS had missing fields.

Nonetheless, the raw metadata requires careful preprocessing before it can be analyzed, particularly for fields like affiliations and venue names. Given the extensive volume of data and the tasks' low requirement for advanced human cognition, we utilized `GPT-4-0125-preview` to review the dataset and create scripts for data correction[2]. This process was supervised and validated by one of the authors.

[2]All the prompts used in this paper are available in Appendix A.

*Table 1.* Paper meta-features considered in this survey.

| ID | Meta feature | Source |
|---|---|---|
| $F_1$ | Authors | WoS |
| $F_2$ | Institutions | WoS |
| $F_3$ | Year | WoS |
| $F_4$ | Title | WoS |
| $F_5$ | Publication venue | WoS |
| $F_6$ | Publication type (Journal or Conference) | WoS |
| $F_7$ | Author keywords | WoS |
| $F_8$ | Web of Science subject category | WoS |
| $F_9$ | Publisher | WoS |
| $F_{10}$ | Number of citations | WoS |
| $F_{11}$ | Number of pages | WoS |
| $F_{12}$ | Country | Eng.† |
| $F_{13}$ | Citation context | SS‡ |

† Eng. implies engineered features. ‡ SS stands for Semantic Scholar.

- For author affiliations, we provided GPT with a list of unique affiliation names, requesting the model to identify and correct potential duplicates that resulted from varied spellings of the same institution or use of abbreviations.
- For publication venues, we asked GPT to inspect the names of conference proceedings, and standardize them to a canonical format (e.g., *Proc. of the 2015 Genetic and Evol. Computation Conference* $\rightarrow$ *GECCO*).

#### 2.3.2 Geographical locations

We also identified the geographical regions of each paper (i.e., $F_{12}$), a feature not directly available from WoS. We engineered this feature by extracting the country name from the author addresses provided by WoS. A document is attributed to a country if at least one author reports an affiliation within that country. Consequently, a paper with multiple authors may be associated with multiple countries.

#### 2.3.3 Citation context

In addition, to analyze the authors' intent when citing MOEA/D, we extracted citation contexts from each paper ($F_{13}$), using the Semantic Scholar `S2ORC` dataset [26]. Each citation context consists of three parts:

- Citation statement: e.g., the sentence containing the in-text citation from the citing article.
- Citation context: e.g., the sentences before and after the citation statement.
- Location of the citation in the citing article: e.g., Introduction, Methods, Results.

Below is an exemplar citation statement. Note that such statements are available only for papers that are open-access or from publishers partnered with Semantic Scholar. This accounts for 3,561 papers in our dataset. Given that a single paper may cite MOEA/D multiple times, this results in a total of 4,675 citation statements.

**A Sample Citation Statement**

"*These algorithms had applied into benchmark problems. Among all these variants only MOEA/D [1], MOEA/D-DD [4], MOEA/D-DU [8], MOEA/D-UR [16], and MOEA/D-URAW [17] had given the comparative results. For this reason, only the results belonging to these 5 algorithms has been reported on the paper ...*" — Section: Results

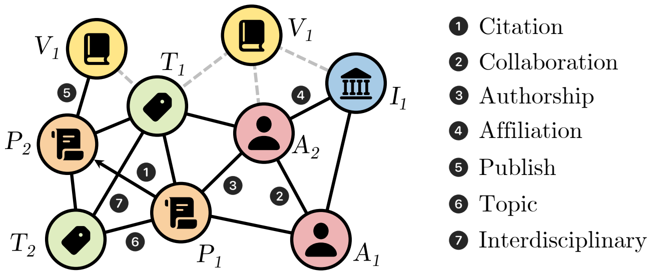


*Figure 2.* An illustration of the MOEA/D knowledge graph, containing 5 types of entities and 7 types of relationships.

## 2.4 MOEA/D Knowledge Graph Construction

We then synthesized all metadata into a bibliographic knowledge graph, as illustrated in Fig. 2. Unlike traditional co-author or citation networks that focus solely on relationships between a single type of entity, our approach constructs a *heterogeneous* network. This network encompasses 5 types of entities, represented as vertices. For simplicity, we use initial capital letters to denote these entity types, and $n, m, q, r, k$ to denote the total number of entities for each type, respectively:

1. Authors, $\mathcal{A} = \{a_i\}_{1 \leq i \leq n}$.
2. Venues, $\mathcal{V} = \{v_i\}_{1 \leq i \leq m}$.
3. Papers, $\mathcal{P} = \{p_i\}_{1 \leq i \leq q}$.
4. Institutions, $\mathcal{I} = \{i_i\}_{1 \leq i \leq r}$.
5. Topics, $\mathcal{T} = \{t_i\}_{1 \leq i \leq k}$.

We extracted the distinct authors, venues, and institutions using the paper metadata outlined in Table 1. At this stage, we utilized WoS subject categories ($F_8$) to represent topics. Further refinement and exploration of topics through advanced modeling techniques are discussed in Section 4. This process resulted in a comprehensive set of entities: $n = 10,532$ authors, $m = 432$ venues, $q = 5,404$ papers, $r = 1,661$ institutions, and $k = 98$ topics, representing a broad and diverse research landscape. Each entity count reflects a careful curation to ensure accuracy and relevance in our knowledge graph.

We then established 7 types of relationships between these entities to elucidate the intricate interconnections within the MOEA/D research community.

1. Citation relations (paper $\Rightarrow$ paper), $E_{PP} = \{e_{ij} | p_i, p_j \in \mathcal{P}\}$. We use `Semantic Scholar API` to extract the citation data. Citation relations in the network are directed, starting from the cited paper to the citing paper.
2. Collaboration relationships (author $\Leftrightarrow$ author), $E_{AA} = \{e_{ij} | a_i, a_j \in \mathcal{A}\}$. We say that two authors are connected if they have co-authored at least one paper. The strength of this link is the number of co-authored papers, weighted by the inverse number of the authors in each paper [27]:

$$S_{AA}(a_i, a_j) = \sum_{p \in \mathcal{P}_{a_{ij}}} \frac{1}{n_p}, \tag{1}$$

   where $\mathcal{P}_{a_{ij}}$ is the set of papers co-authored by $a_i$ and $a_j$, and $n_p$ is the number of authors in paper $p$.
3. Authorships (paper $\Rightarrow$ author), i.e., $E_{PA} = \{e_{ij} | p_i \in \mathcal{P}, a_j \in \mathcal{A}\}$.
4. Affiliation relationships (author $\Rightarrow$ institution), i.e., $E_{AI} = \{e_{ij} | a_i \in \mathcal{A}, i_j \in \mathcal{I}\}$.
5. Publication relationships (paper $\Rightarrow$ venue), i.e., $E_{PV} = \{e_{ij} | p_i \in \mathcal{P}, v_j \in \mathcal{V}\}$.
6. Topic relationships (paper $\Rightarrow$ topic), i.e., $E_{PT} = \{e_{ij} | p_i \in \mathcal{P}, t_j \in \mathcal{T}\}$. Each paper can belongs to multiple topics.

7. Interdisciplinary relations (topic $\Leftrightarrow$ topic), i.e., $E_{PT} = \{e_{ij} | p_i \in \mathcal{P}, t_j \in \mathcal{T}\}$. We say two topics are connected if they co-occur in the same paper, and the strength of this link is defined as the Jaccard similarity coefficient of the set of papers that contain each topic:

$$S_{TT}(t_i, t_j) = \frac{|\mathcal{P}_{t_i} \cap \mathcal{P}_{t_j}|}{|\mathcal{P}_{t_i} \cup \mathcal{P}_{t_j}|}, \tag{2}$$

where $\mathcal{P}_{t_i} = \{p \in \mathcal{P} | t_i \in p\}$.

Finally, we construct the bibliographic knowledge graph formally as a triplet $G = \langle V, E, F \rangle$, where

- $V = V_A \bigcup V_V \bigcup V_P \bigcup V_I \bigcup V_T$ represents the set of vertices, i.e., authors, venues, papers, institutions, and topics.
- $E = \langle E_{PP}, E_{AA}, \cdots \rangle$ denotes the set of edges, indicating various types of relationships among these entities.
- $F = \langle F_1, F_2, \dots \rangle$ encompasses the attributes of each vertex, such as 'publication year' for papers, 'paper counts' for authors, and 'country' for institutions, among others.

This knowledge graph is *heterogeneous*, reflecting the diversity of entities and relationships; *directed*, indicating the directionality of relationships; and *attributed*, capturing the characteristics of each entity. It essentially comprises all the information necessary for conducting the comprehensive analysis that unfolds in the remaining sections of this survey.

# 3 General Data Analysis

In this section, we provide a general overview of the knowledge graph constructed in Section 2, while more advanced analysis based on network mining and topic modeling will be presented in the follow-up sections.

### 3.0.1 Publication trends

The histogram in Fig. 3(A) depicts the annual distribution of publications related to MOEA/D. The first follow-up works on MOEA/D appeared in 2008, with the number of publications experiencing a significant upward trend thereafter, surpassing 100 publications in 2013, reaching 500 by 2019, and nearing 800 in 2023. The data exhibit an average annual growth rate of 40.50%. Note that data for 2024 are not included in this analysis, as complete statistics are not yet available at the time of this writing. Additionally, the cumulative number of publications, also depicted in Fig. 3(A), closely follows a quadratic growth trend, evidenced by a coefficient of determination $R^2 = 0.999$. This suggests that the MOEA/D research community can expect the continuation of a robust growth trend, with an increasing number of papers anticipated to address further challenges in the field.

**Remark 3**

The publication date considered in this analysis refers to the official date when the work is listed on the publisher's website with an assigned digital object identifier. However, it is not uncommon for many studies to be available as citable pre-prints before their formal publication. As a result, discrepancies may arise between different data sources.

### 3.0.2 Researcher involvement

Fig. 3(B) shows the annual count of authors involved in MOEA/D publications and its cumulative distribution. These plots echo the trends observed in Fig. 3(A). Notably, since 2017, over $1,000$ researchers have contributed to MOEA/D research annually, and the total number of researchers exceeded $10,000$ earlier 2024. The cumulative distribution of researchers follows a quadratic trend ($R^2 = 0.998$), suggesting an ongoing increase in the number of researchers joining this community.

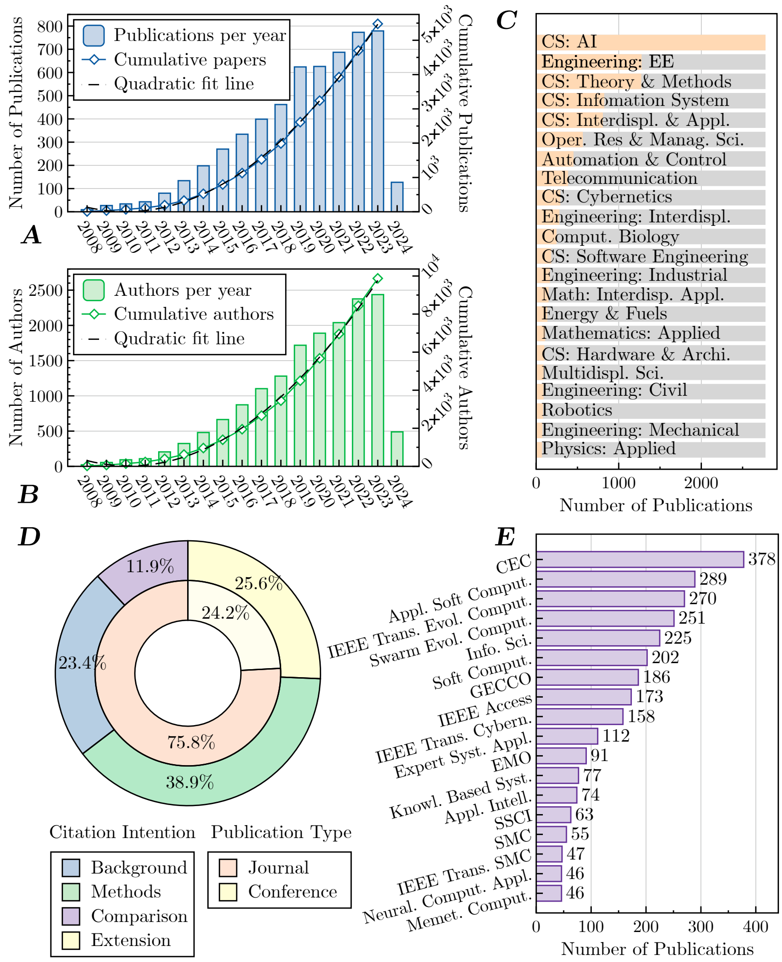


*Figure 3.* General information of surveyed MOEA/D literature. **(A)** Number of pulications per year and its cumulative distribution. **(B)** Number of authors per year and its cumulative distribution. **(C)** Frequency of top-20 subject categories. **(D)** Ring chart of pulication type (inner) and citation intention (outer). **(E)** Number of publications of the top-20 popular venues.

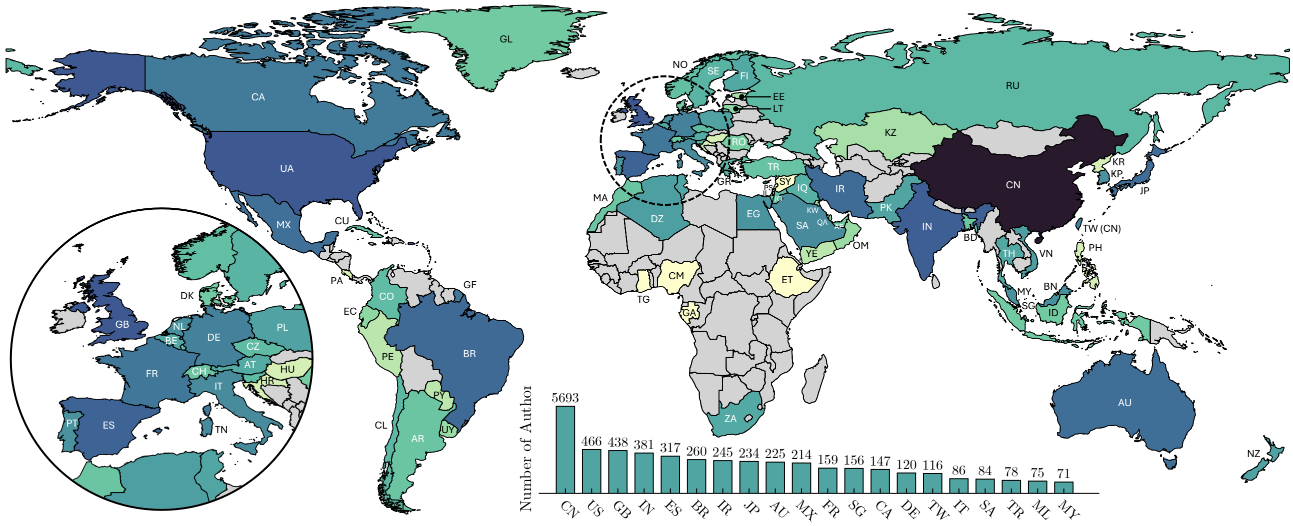


*Figure 4.* **(Top)** Geographic distribution of MOEA/D researchers. **(Bottom)** The number of researchers in the 20 most active regions.

### 3.0.3 Publication venues

Fig. 3(E) presents the most prevalent venues for MOEA/D publications, while the inner ring of Fig. 3(D) illustrates the distribution by type. We can see that journals are the primary publication

*Figure 5.* Placeholder for disruptiveness plot.

venue, accounting for 76% of all publications. Among these, Appl. Soft. Comput, IEEE Trans. Evol. Comput., and Swarm. Evol. Comput. have the highest number of publications. Conferences and symposiums represent the remaining 24%, led by CEC, GECCO, and EMO. Although Evol. Comput. (26 papers) and PPSN (39 papers) are well recognized within the community, their publication counts are relatively low, placing them outside the top ranks.

Moreover, the diversity of venues extends beyond the evolutionary computation community into other computational intelligence domains such as IEEE Trans. Fuzzy Syst., IEEE Trans. Intell. Transp. Syst., IEEE Trans. Geosci. Remote Sens., and IEEE Trans. Neural Netw. Learn. Syst.. These indicate an interdisciplinary research landscape of MOEA/D. To have a better visualization of domain diversity, Fig. 3(C) categorizes publications across various subject areas within the WoS. While this provides a coarse-grained categorization, a more elaborated topic modeling is discussed in Section 4.

#### 3.0.4 Geographical regions

We visually mapped the geographical distribution of MOEA/D research authors in Fig. 4. From this diagram, we observe that research on MOEA/D is particularly active in China, which accounts for over half of the total researchers in this field. Following China, the United States, the United Kingdom, India, and Spain are among the top contributors to MOEA/D research. This pattern underscores not only the dominance of specific nations but also a broad, global engagement with MOEA/D research, spanning 82 countries across all the major continents.

#### 3.0.5 Citation intents

By analyzing citation statements ($F_{13}$) as described in Section 2, we were able to infer the underlying intentions behind each citation of MOEA/D, contingent on the availability of citation statements. While the classification of citation intents has been the subject of previous studies (e.g., [28, 29]), it is only with the recent advancements in LLMs that we have seen a notable increase in classification performance [30]. Inspired by this, we prompted `GPT-4` to annotate the 4,675 citation statements we gathered into four distinct categories, which are illustrated in the outer ring of Fig. 3:

- Background: Approximately a quarter of the citation statements mentioned MOEA/D in the background section, providing historical context, justifying significance, or offering other information directly related to [1].
- Method: Nearly 40% of the citations referred to MOEA/D as their main methodology. These papers often focus on applying EMO in other disciplines.
- Extension: Another 26% of the citations claim to extend the original MOEA/D framework, introducing new features, capabilities, or adaptations.
- Comparison: The remaining 12% of the citations conducted performance comparisons with MOEA/D, often to benchmark new methods or improvements.

## 4 Topic Modeling

To uncover prevalent research trends within the MOEA/D domain from our constructed knowledge graph, we utilized topic modeling [4], a technique from the NLP domain, to cluster papers based on textual analysis of their titles and abstracts. These sections typically offer concise summaries of a paper's focus and contributions, making them ideal for extracting thematic insights. Topic modeling on academic papers has been effectively used across diverse fields, including information systems [31],

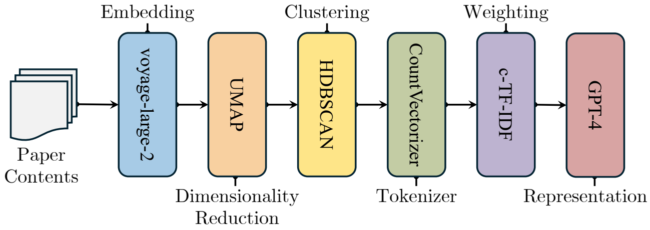


*Figure 6.* A high-level overview of the `BERTopic` framework with our adopted implementations.

computer systems [15], transportation research [16], software engineering [32], healthcare [33]. However, existing works predominantly applied probability-based topic models like non-negative matrix factorization (NMF) [34] and latent Dirichlet allocation (LDA) [35], which fall short of capturing the nuanced semantic similarities between papers, and suffer from poor scalability when handling large text corpora [36].

In recent years, bidirectional encoder representations from transformers (BERT) [37] and its variants (e.g., [38–40]) have achieved significant advancements in generating vector representations of texts which incorporate their inherent semantic meanings. Leveraging these contextual representations, more advanced topic models have been developed by examining the distances between the paper's vector representations in the embedding space [41–43]. They offer a leap in topic coherence, diversity, and scalability compared to traditional models [36].

In this work, we adopted `BERTopic` [41], a well-established embedding-based topic modeling framework, combined with the state-of-the-art text embedding models and the power of LLMs, as well as expert consultations, to offer the first detailed scrutiny of the enormous research landscape of MOEA/D.

## 4.1 BERTopic Framework Implementation

A high-level overview of the `BERTopic` framework is shown in Fig. 6, which consists of five building blocks connected in a sequential workflow, where each can be implemented with different methods.

### 4.1.1 Input documents

While existing practices typically use either paper titles or abstracts as input documents for the topic models, here we combine *i*): titles, *ii*): abstracts, and *iii*): keywords, to incorporate more comprehensive information. Additionally, it has been recently shown that *instruction fine-tuning* by providing further descriptions on the type, objective, and domain of the task can help the model's reasoning and tailor the generated embeddings to specific downstream tasks [44]. We thus input the source text corpora for each paper as follows, taking [45] as an example:

**Textual Information for Embedding**

"*represent the paper on multi-objective optimization for topic modeling:* ***title:*** *'an evolutionary many-objective optimization algorithm based on dominance and decomposition.'* ***keywords:*** *'decomposition, evolutionary computation, many-objective optimization, pareto optimality, constrained optimi ...'* ***abstract:*** *'achieving balance between convergence and diversity is a key issue in evolutionary multiobjective optimization ...'.*"

### 4.1.2 Document embedding

The first step in the `BERTopic` framework is to embed the input documents to create representations in vector space, where documents of similar topics will lie close to each other. For this purpose, we adopt `voyage-2-large` model from VOYAGE AI, which can generate better embeddings compared to traditional models under the Sentence BERT (SBERT) framework [46] or other commercial models like the `text-embedding-3-large` from OpenAI on various sentence embedding tasks [47]. This will then enable BERTopic to perceive more nuanced details in the paper contents and generate more accurate topics.

#### 4.1.3 Dimensionality reduction

The embeddings generated by voyage-2-large are in a $1,565$-D space, rendering the concept of spatial locality ill-defined. In other words, due to the curse of dimensionality [48], distance measures will become approximately the same, posing challenges for clustering algorithms. To address this, BERTopic first compresses the obtained embeddings to a low-dimensional space via dimensionality reduction. We adopt UMAP [49] as it is more computationally efficient than other methods like $t$-SNE [50].

#### 4.1.4 Document clustering

We then identify topics by clustering the reduced embeddings using HDBSCAN [51], which can find clusters of varying densities via a hierarchical clustering approach. It can also handle noises as outliers and does not force unrelated papers to be assigned to any cluster, thereby improving topic representations.

#### 4.1.5 Topic representation

After obtaining clusters of paper, BERTopic then generates a topic in natural language using a class-based variation of TF-IDF [52]. Additionally, as we will see later, for many interdisciplinary topics (e.g., hyperspectral imaging), it is still challenging for general readers to understand the topic even with TF-IDF keywords. In light of this, we further employed GPT-4 to generate a concise and easily understandable summary for each topic by providing it with TF-IDF representation, paper titles, and examplar abstracts.

In practice, we iteratively fine-tuned the instructions in Step *1)* and conducted model selection and hyperparameter tuning for Steps *2) to 5)*. We also closely collaborated with domain experts in different fields during this process. Their expertise and feedback were incorporated into the framework via the *zero-shot* modeling capability of BERTopic.

By applying this pipeline, BERTopic eventually identified 83 distinct research topics, with 20% papers identified as outliers. We plot the generated keywords along with their relative frequencies using wordclouds map shown in Figs. 9 and 10. Here, topics are ranked by their sizes, where T0 is the largest one while T82 is the smallest. From the wordclouds, it is clear that each topic represents some coherent area of research, which demonstrates the effectiveness of BERTopic in capturing themes in the MOEA/D literature. GPT summarizations of these topics are available in Appendix B.

Utilizing wordclouds and low-dimensional visualization of the paper embeddings (see Fig. 8), we categorize the MOEA/D research topics into two main types: methodological enhancements and extensions (40 topics), and application-driven research (43 topics). The former encompasses studies focusing on refining and advancing MOEA/D methodologies, while the latter includes works applying the method across various domains. Each type is further subdivided into distinct themes, with the hierarchical structure of these themes illustrated in Fig. 7.

**Remark 4**

Outliers in the context of topic modeling do not necessarily imply the papers are of low quality or irrelevant to our survey. Instead, it may suggest a paper is lying in a mid-region between two or more topics in the semantic space. Although outliers can be reduced or eliminated by adjusting the hyperparameters of the clustering algorithm, this will lead to less coherent topics, and the same is true for using overlapping clustering methods. We thus allow a reasonable amount of outliers in our results.

**Remark 5**

The size of the topics here should be interpreted with caution. One reason for this is the presence of outliers, which implies that we should more focus on the relative sizes of the topics. In addition, since this paper is routed in the EMO field, we tuned BERTopic to zoom in more into the methodological topics, while maintaining only a bird-view of the application topics. This is why T0 and T1 are larger than T2 and T3.

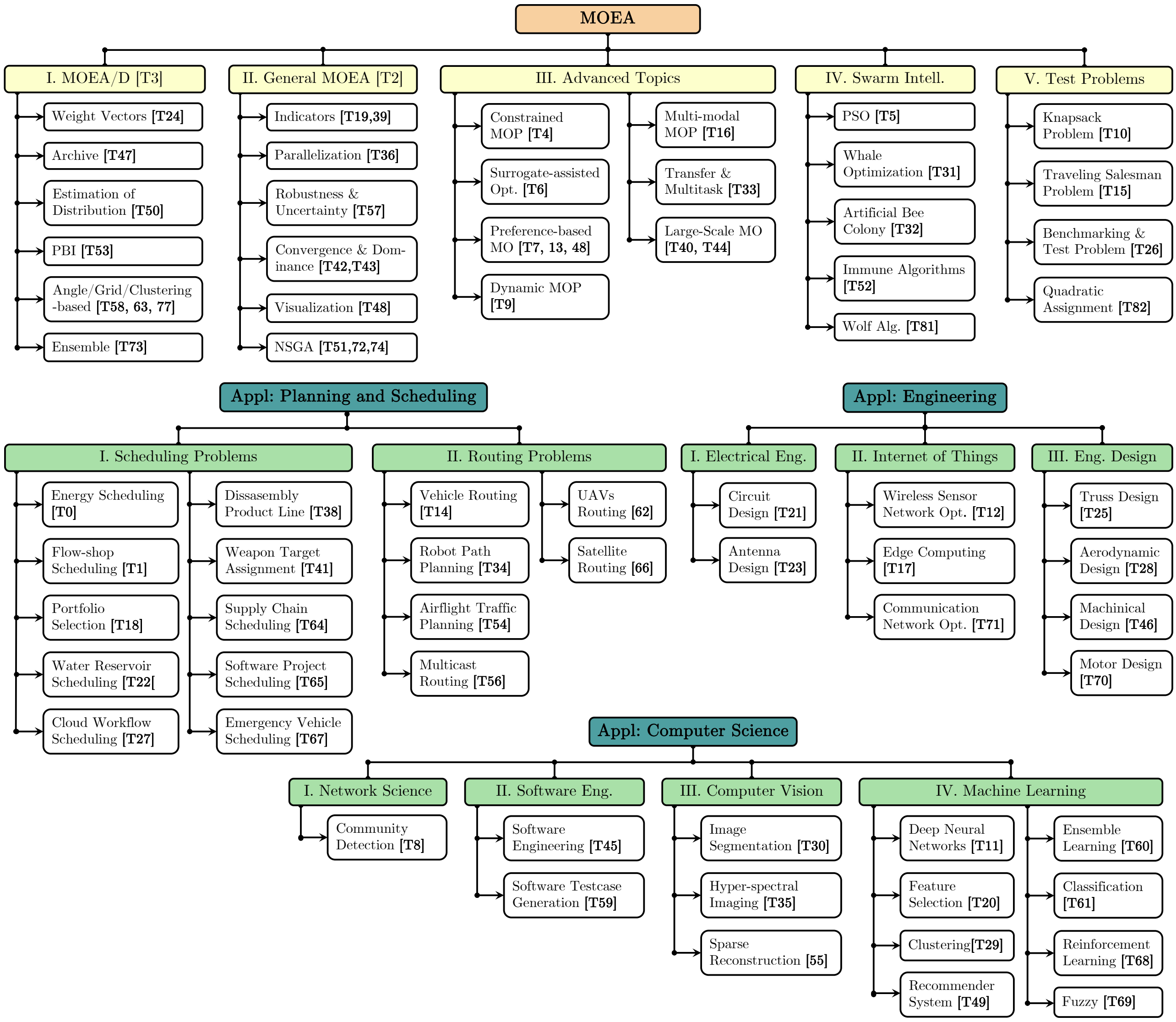


*Figure 7.* Hierarchical structure of the `BERTopic` topics.

## 4.2 Methodological Topics

The methodological topics pertains to the general family of MOEAs, which consists of five themes.

### 4.2.1 MOEA/D

Topics under this research theme have been thoroughly reviewed in Part I of this survey series. Many follow-up works on MOEA/D fall into `T2` [53–55], which is the largest topic in the methodology section. While the `T2` is more like a general one, we noticed that several topics mentioned in Part I formed independent groups, e.g., weight vectors settings (`T24`), archives (`T47`), estimation of distribution methods (`T50`), and penalties (`T53`).

### 4.2.2 General MOEA

This theme covers various topics that are concerned by the broader EMO community, such as performance indicators (`T19` and `T39`), parallelization (`T36`), knee points (`T48`), robust optimization (`T57`), visualization (`T79`), etc. We also identified groups of papers focusing on exploring the dominance relationships (`T43`) as well as seeking balances between convergence and diversity (`T42`). In addition, there are several topics pertaining to NSGA-II [56] and NSGA-III [57]. They often cite MOEA/D for comparison purposes. This consolidates the rationality of our paper selection protocol in Section 2. Note

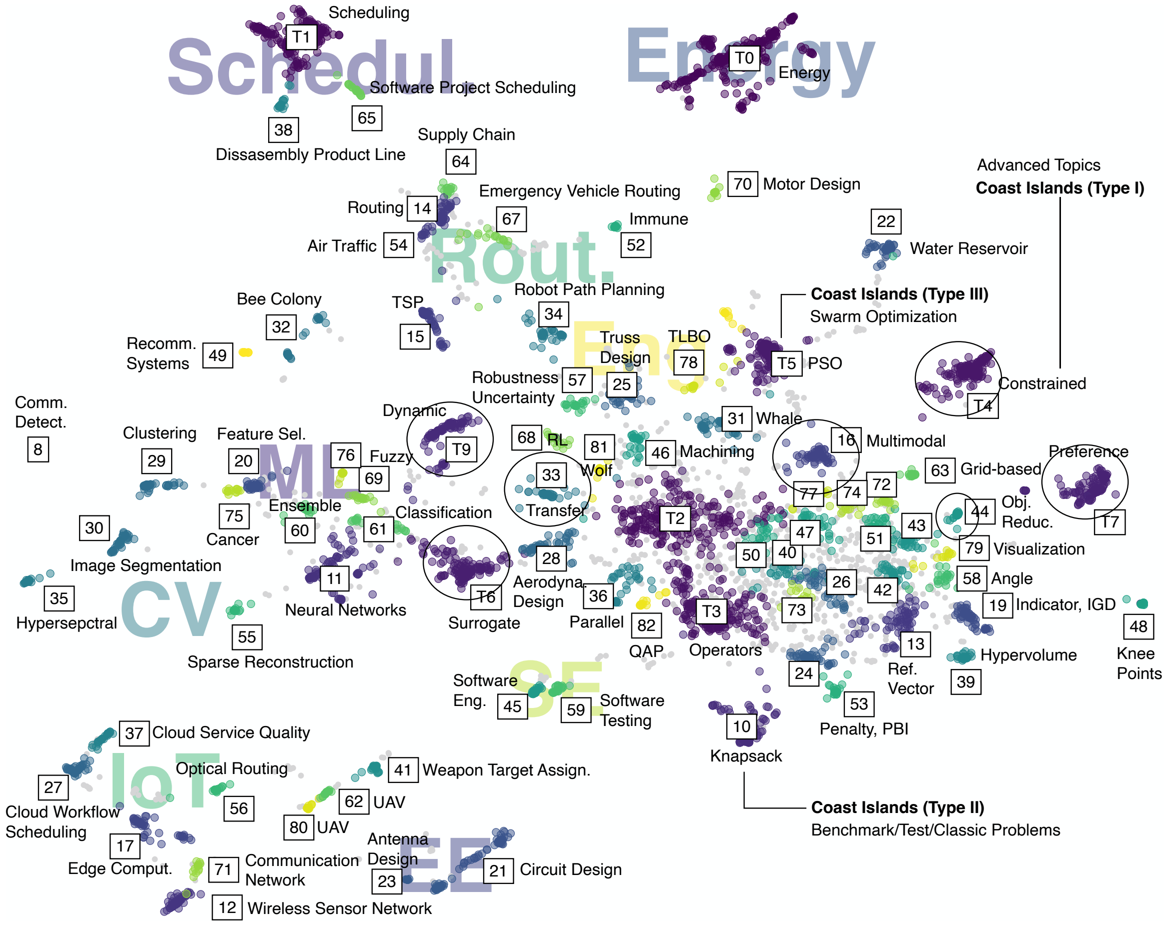


*Figure 8.* Low-dimensional visualization of the MOEA/D literature landscape by projecting the paper embeddings using UMAP. Papers are colored and labeled by `BERTopic` topics. Outliers are shown in light gray. Advanced topics discussed in Part I are circled.

that having seemingly unrelated papers would not introduce issues to topic modeling, as `BERTopic` is able to identify them as isolated topics. In contrast, the inclusion of such papers is beneficial for a bird-view of the whole EMO community and pinpointing the position of MOEA/D. Finally, we make notes about `T3`, the general MOEA topic. It primarily contains survey papers and guidance/notes on MOEAs. Within it there are also software and tools that implement a large number of MOEAs (e.g., `Pymoo` [58] and `PlatEMO` [59])

### 4.2.3 Advanced topics

The next theme incorporates various advanced contents covered in Part I. Notably, constrained MO (`T4`), surrogate modeling (`T6`), preference-based and interactive MO (`T7` and `T13`), dynamic MO (`T9`) are among the most popular methodological topics. In addition, we also noticed emerging interests including multi-modal MO (`T16`), transfer learning and multi-tasking (`T33`), as well as large-scale MO (`T40`) and objective reduction (`T44`). For detailed discussions on these topics, we refer the readers to Part I.

### 4.2.4 Swarm intelligence

Another line of research that has been gaining traction is the swarm-based methods, which has also been covered in Part I. In particular, this line is led by particle swarm optimization (`T5`, e.g., [60–62]). Another related topic includes whale optimization (`T31`, e.g., [63]), artificial ant/bee colony (`T32`, e.g., [64–67]), immune algorithms (`T52`, e.g., [68]), and grey wolf optimizers (`T81`, e.g., [69]).

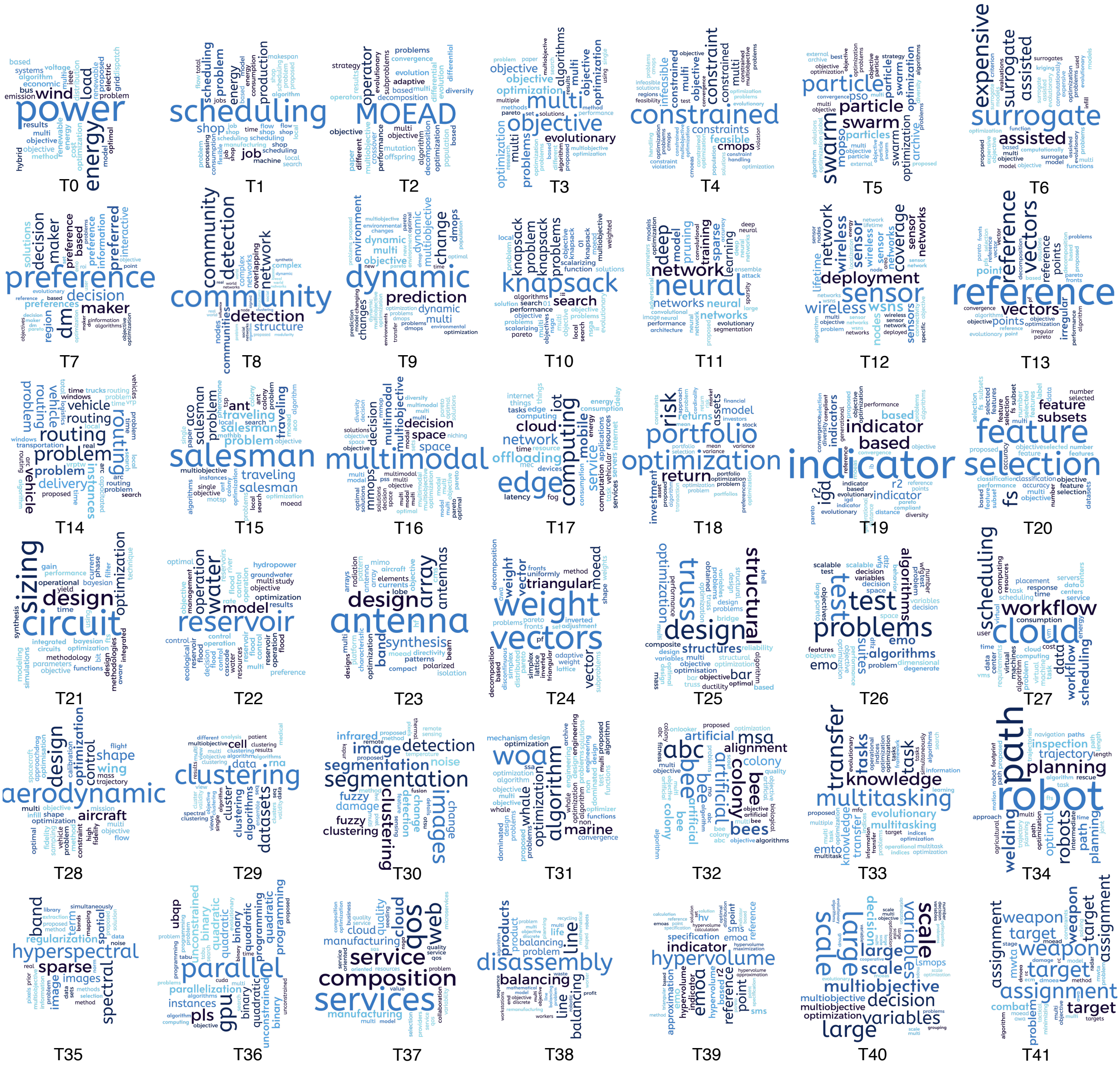


*Figure 9.* Wordcloud map for the `BERTopic` topics (part 1, word size indicates frequency).

#### 4.2.5 Benchmark test problems

The final theme in the methodological section is about classic MOPs. Among them, multi-objective knapsack problem (`T10`, e.g., [70]) is the most popular one, followed by multi-objective traveling salesman problems (`T15`, e.g., [71]). Then, we have a dedicated topic for multi-objective benchmarking and test problems (`T26`, e.g., [72]). There are also works on multi-objective quadratic assignment problems (`T82`, e.g., [73]).

### 4.3 Applications

In addition to the methodological advancements, topic modeling uncovers a diverse spectrum of MOEA/D applications, which can be grouped into three thematic realms.

#### 4.3.1 Operations research

A huge portion of MOEA/D applications are in essence planning, scheduling, or routing problems that are concerned in operations research (OR), while these topics themselves may stem from different

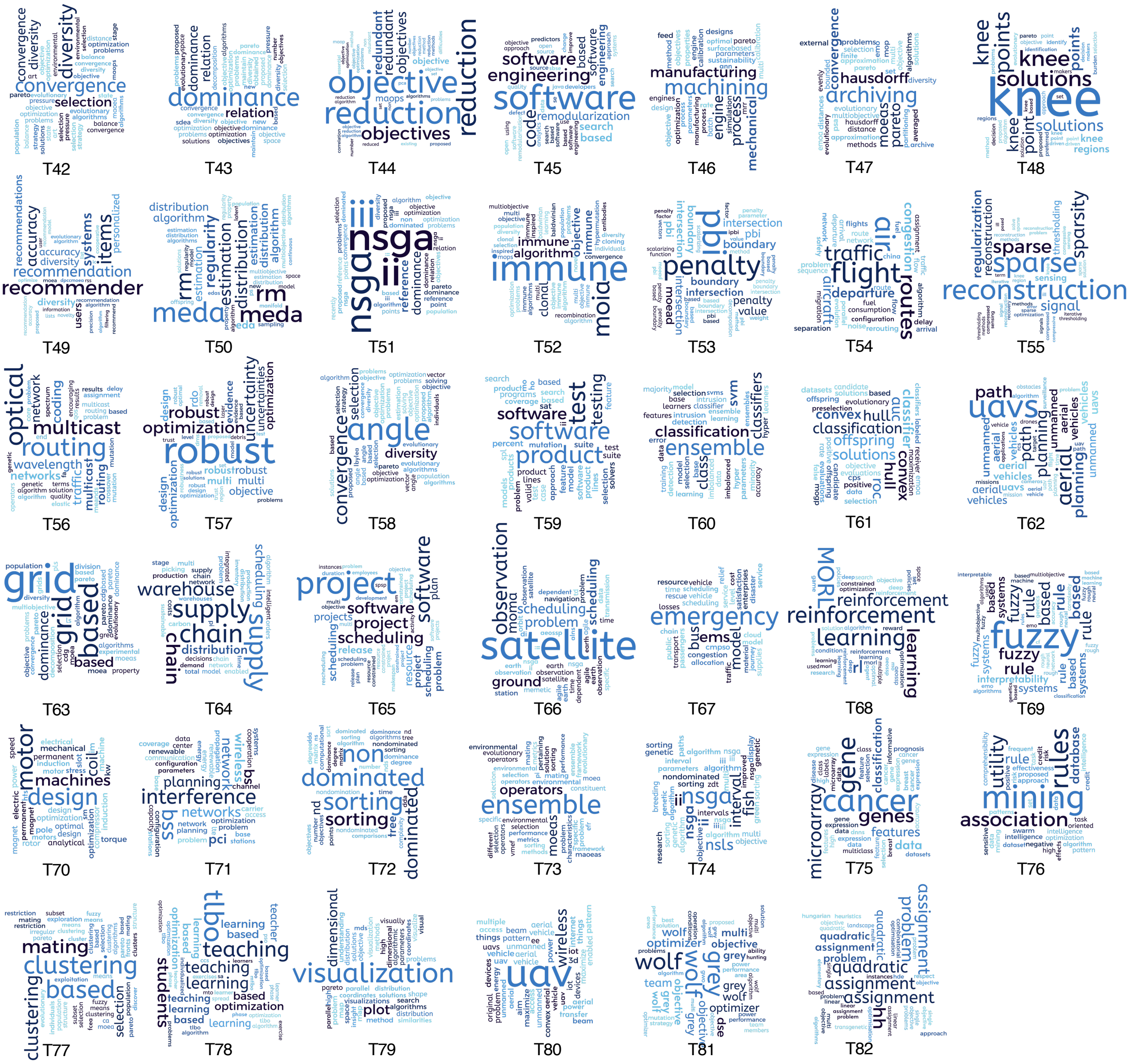


*Figure 10.* Wordcloud map for the BERTopic topics (part 2, word size indicates frequency).

domains. For instance, ▶ the *planning & scheduling* topics are associated with various engineering problems (e.g., [74–76]) like flow-shop (T1) and supply chain (T64) scheduling, and there are also portfolio selection problems from the finance domain (T18, e.g., [77]), software project scheduling from the software engineering domain (T65, e.g., [78]), as well as weapon target assignment problem from military OR (T41, e.g., [79]). On the other hand, ▶ the set of *routing problems*, e.g., [80], involve different types of vehicles, from robotics (T34) and cars (T14), to UAVs (T62) and aircrafts (T54), and even satellites (T66).

### 4.3.2 Engineering

MOEA/D has also been widely applied in various real-world engineering scenarios. For instance, ▶ assisting the design of analog circuit (T21) and antenna (T23) in electronics engineering (e.g., [81]), ▶ optimizing wireless sensors, devices (T12, T17) and communication networks (T56, T71) in communication engineering (e.g., [82–84]), as well as ▶ facilitating various types of engineering shape design problems (T25, T28, T46, T70, e.g., [85–87]).

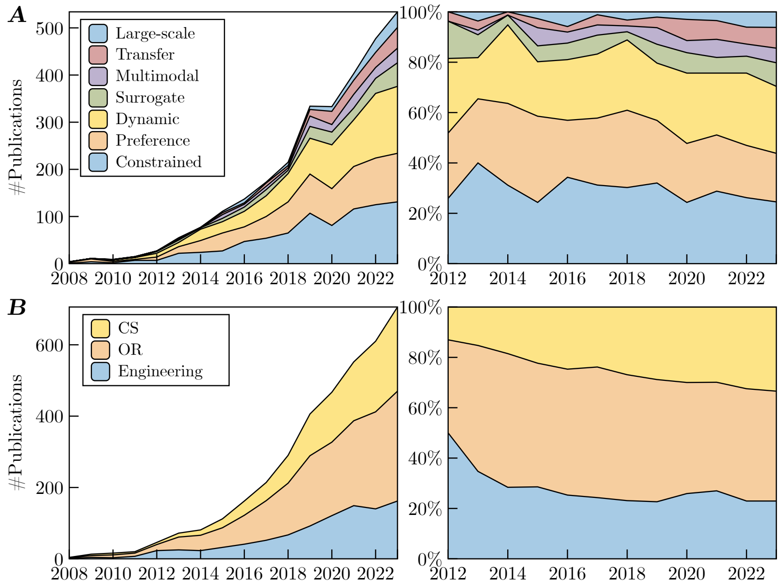


*Figure 11.* **(Left)** Number of publications per year and **(Right)** relative percentages for **(A)** the 7 topics on MOs variants, **(B)** the 3 application domains. For the right panel, we truncated the time range to 2012-2023 as the number of publications before 2012 is relatively small.

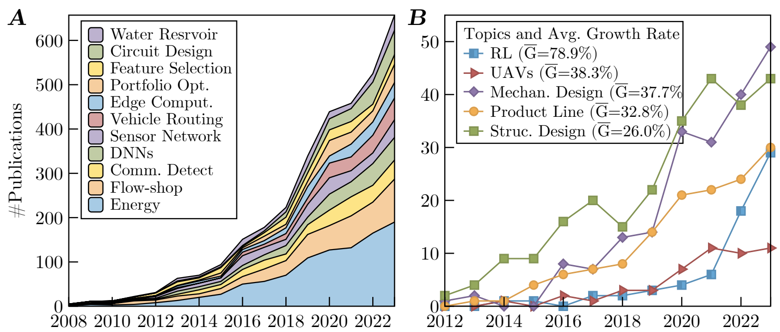


*Figure 12.* Number of publications per year for **(A)** top-10 popular applications topics and **(B)** top-5 emerging application topics.

### 4.3.3 Computer science

In the recent decade, with the advancements in both hardware and algorithms, the computer science (CS) domain has also seen a surge in MOEA/D applications. This is especially true for ML, where the most prominent examples are multi-objective neural architecture search (`T11`, e.g., [88]) and multi-objective feature selection (`T20`, e.g., [89]). In addition, for other ML tasks like clustering (`T29`, e.g., [90]), recommendation (`T49`, e.g., [91]), and reinforcement learning (RL, `T68`, e.g., [92]), considering multiple conflicting objectives has also received growing attention.

## 4.4 Topics distribution over time

In this study, our interest is not confined to the spatial distribution of current research topics. We are also interested in investigating the temporal dynamics of the research over time. While `BERTopic` offers a method for dynamic topic modeling, challenges such as inaccuracies from unclassified papers (i.e., outliers) and the model's limitations in addressing overlapping topics necessitate a tailored approach. To overcome these issues, we manually crafted query strings based on paper abstracts to compile an extended set of papers for each topic identified by `BERTopic`. This adjustment allows for papers to be assigned to multiple topics, enhancing the granularity of our analysis. Then, we employed the method in [93] to model the temporal distribution of topics, with a particular focus on the seven advanced variants of MOs and applications of MOEA/D.

### 4.4.1 Advanced research topics

Fig. 11 depicts the evolution of the seven advanced topics from 2008 to 2023, revealing a growth trend consistent with the overall increase in publications, as previously shown in Fig. 3. Notably, topics related to constrained MOPs (`T4`), preference-based MOPs (`T7`, `T13`, `T48`), and dynamic MOPs (`T9`) are particularly prominent (Fig. 11A), collectively accounting for approximately 400 papers published in 2023. Even after adjusting for overlap, these topics contribute to 319 unique publications. Considering the total publication count of about 800 in 2023, nearly 40% of the papers touch on at least one of these topics, highlighting their significant interest within the community. Nonetheless, the relative share of these three dominant topics has seen a decline over time, despite their growth in their absolute numbers (Fig. 11(A), right). This shift can be attributed to the emergence and growth of other topics such as surrogate-assisted MO (`T6`), multi-modal MO (`T16`), multi-task MO (`T33`), and large-scale MO (`T40`, `T44`), indicating a diversification in research interests.

### 4.4.2 General trend on applications

We then take a look at Fig. 11(B), which shows the trends for 3 general MOEA/D application domains. We found that while the number of publications all underwent a steady growth, the pace has been faster for the computer science domain ever since 2012, which is in line with the rapid development of ML in the past decade. This surge is contrasted by a proportional decline in engineering-focused applications, although the OR domain has sustained a consistent proportion of publications. However, we note that as discussed before, many topics under OR actually stem from engineering domains. This nuance suggests that the observed shift may not truly reflect a departure from traditional engineering applications but rather highlight the expanding diversity and interdisciplinary nature of MOEA/D applications.

### 4.4.3 Top-10 application topics

Fig. 12(A) illustrates the evolution of the top-10 application topics of MOEA/D, among which energy (`T0`), flow-shop shceduling (`T1`), and community detection (`T8`) are the most prevalent ones. In particular, we can see that applications in the energy sector have increased dramatically in the past decade, which can be regarded as part of the global efforts to advance sustainable development goals (SDGs) of the United Nations. Other popular application topics include neural networks design (`T11`), wireless sensor network optimization (`T12`), vehicle routing (`T14`), edge computing (`T17`), portfolio selection (`T18`), feature selection (`T20`), analog circuits design (`T21`), and resorvior management (`T22`).

### 4.4.4 Emerging application topics

We also identified 5 most emerging application topics of MOEA/D, as ranked by their annual growth rate in publications since 2018. Note that several topics appeared in Fig. 12(A) also exhibit high growth rates, but here we adhere to ones that are relatively new to the community. From the results in Fig. 12(B), we can see that RL (`T68`) has been the most rapidly growing application of MOEA/D, whose publications have increased by nearly $4\times$ in the past two years. Further, with the maturity and commercialization of UAV technologies, applications of MOEA/D in this domain (`T62`) have also gained attention since 2020. In addition, for some traditional domains, e.g., structral (`T25`) and mechanical (`T46`) design, as well as product line scheduling structral (`T38`), we also observed a significant increase in the number of publications in the past several years.

Finally, an interesting aspect of the temporal distribution is the noticeable 'pause' or decline in the growth trends of topics illustrated in Fig. 11(A) and Fig. 12(A) during 2020, likely attributed to the disruptions caused by the COVID-19 pandemic. Despite this setback, the growth trends showed a robust recovery starting from 2022 onwards. Interestingly, this pandemic-induced impact appears less pronounced in the broader trends shown in Fig. 11(B). This resilience may stem from the inclusion of a wide array of topics, which contributes to the overall resilience against fluctuations in specific areas of research.

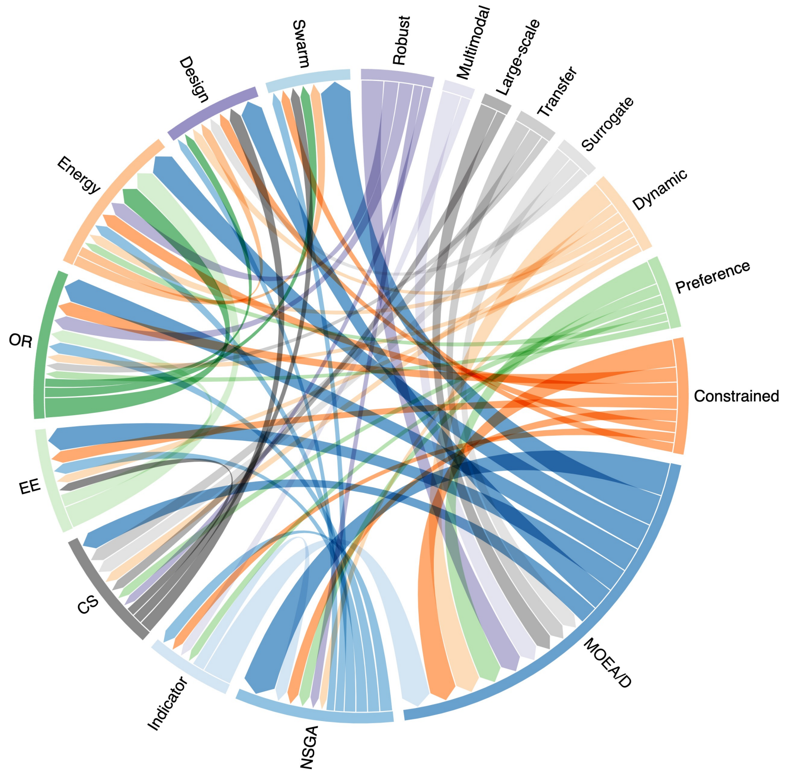


*Figure 13.* Chord diagram showing the linkages between different sectors of research themes in the MOEA/D landscape. This is determined via keyword co-occurrence in paper abstracts. A thicker chord indicates a stronger linkage between two topics. Direction of the linkages is only for enhanced readability. Linkages with strengths below $\epsilon = 0.15$ (equation (2)) are masked.

### 4.5 Topic Linkages

The extended overlapping paper set for topics also provides a useful lens to analyze the linkages among major themes of research in the MOEA/D landscape following equation (2). We present the results as a chord diagram in Fig. 13. Starting with the general MOEA/D topic, we observe strong connections with nearly all other topics, encompassing both methodological studies and real-world applications. This demonstrates the broad impact of MOEA/D across various areas. Moving counter-clockwise along the circle are the 7 advanced variants of MO, showing their primary associations with applications in addition to their connections with the core MOEA/D theme.

Specifically, constrained MO is the most prevalent among the 7 variants, drawing significant attention across all major application domains. Preference-based MO shows a strong presence particularly in CS, OR, and power engineering. Note that the absence of preference incorporation in other domains simply indicates its relative emphasis falls below a certain threshold, $\epsilon = 0.15$. Similarly, dynamic MO demonstrates robust linkages with all application fields, aligning with our earlier observation of a dominant and increasing interest in these three topics. Regarding the remaining four MO variants, their connections with CS are notably strong, except for multi-modal MO. Furthermore, multi-task MO is also prominent in OR, and many engineering design tasks adopt surrogate modeling.

We also observed the intertwining between different application domains. For instance, many studies within power engineering essentially tackle planning and scheduling problems traditionally addressed in OR. There is also close relationship between power engineering and the broader field of electrical engineering (EE), highlighted by a prominent chord connecting the two topics. In addition to these, topics on NSGA and swarm optimizers are also shown to be linked with many topics in the diagram, but with a much weaker strength, as our survey is focused on MOEA/D. We also explore topics related to performance indicators in MO, which mainly interacts with methodological themes. Additionally, the subject of robustness and uncertainty in MO has attracted interest across both

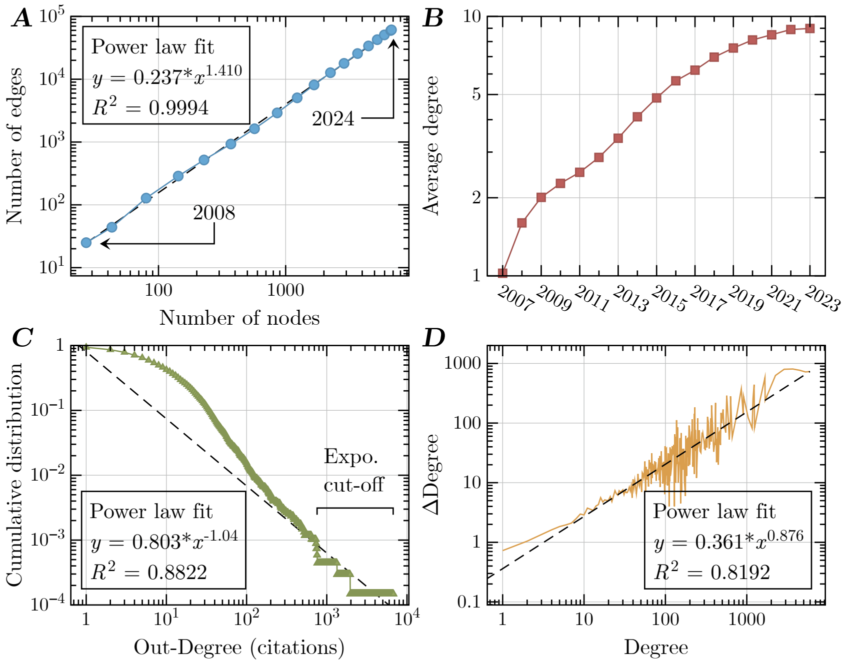


*Figure 14.* General pattterns for MOEA/D citation networks. **(A)** The growth of the network nodes and edges from 2008 to 2024. **(B)** The average degree of the network each year. **(C)** Cumulative distribution of the number of citations of each paper within the network. **(D)** The number of new citations of a paper per year as a function of the citation it has already collected.

methodological and application domains.

# 5 Citation Network Analysis

While in Section 4, topic modeling based on advanced NLP and ML techniques has provided one-of-a-kind bird-view of the entire research landscape of MOEA/D, it treated each paper as an isolated entity. However, scientific publications are inherently interconnected through citation networks, which serve as channels for knowledge accumulation and dissemination. Citation network analysis [94] has been extensively studied to understand the themes of research, to trace the evolution of fields, and to identify or forecast influential works [18, 95]. In this section, we conduct graph mining of the citation relationships in our constructed knowledge graph to illuminate the MOEA/D research landscape from four different perspectives.

## 5.1 Network Evolution

The MOEA/D citation network has experienced significant growth over the past decade. As shown in Fig. 14(A), the network's expansion adheres to the *densification power law* [96], with an exponent $\alpha = 1.41$, indicating an acceleration beyond linear growth ($\alpha = 1$). This results in an increasing average out-degree (expected citations within the network) over time as shown in Fig. 14(B), making the network denser.

Despite this overall growth in network density, not all papers are equally likely to get new citations. From Fig. 14(C), we find that the citation[3] distribution among the network is highly skewed, which also follows power law, with exponent $\alpha = 1.04$. In particular, while the expected paper citation grows over time, only approximately 6% papers garnered $> 10$ citations as of March 2024, merely 1% exceeded 100 citations, and only a handful of works received more than $1,000$ citations. Similar patterns, known as *scale-free* networks [97], have also been observed for citation networks of other

[3]The citation index discussed throughout this section is defined as the out-degree of each node in the constructed MOEA/D citation network. The value can thereby be lower than total citations of a paper.

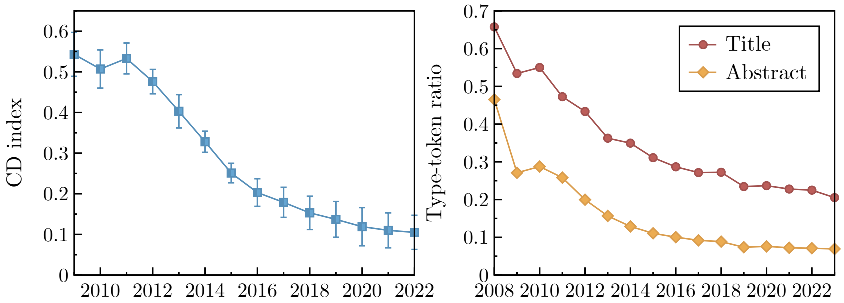


*Figure 15.* Paper disruptiveness for MOEA/D literature across time as measured by **(Left)** the CD index **(Right)** type-token ratio in paper titles and abstracts. Note that as the calculation of CD index requires information regarding both ancestor and successor papers, the data ranges from 2009 to 2022.

domains [98]. We contend that there are the following three factors related to the polarization in the paper citations.

**Preferential citations.** One well-known model that elucidates the degree accumulation mechanism in scale-free networks is the *preferential attachment* [97]. This concept suggests that nodes newly added to the network tend to connect with those already having a higher degree. Applied to the citation network, this phenomenon is depicted in Fig. 14(D), where the annual increase in citations for a paper is plotted against its existing citation count. It is evident that researchers are prone to cite well-recognized articles over those less-cited ones, with a power law exponent of $\alpha = 0.876$. This bias towards more cited works underpins the scale-free structure of the MOEA/D citation network, exemplifying the 'rich-get-richer' dynamic.

**Knowledge obsolescence.** While preferential attachment governs the creation of new citations, this growth dynamics can be counterbalanced by other factors. One such example is known as knowledge obsolescence [99], which suggests a tendency towards obsolescence for old knowledge. As a result, the citation growth of the top-cited papers can slow down over time, and the distribution exhibits an exponential cut-off [100], as can be observed in Fig. 14(C) for citation $> 1,000$.

**Competition.** Another factor that constrains citations is the competition among papers on similar topics according to a *fitness* score [101]. This fitness reflects a paper's perceived relevance and attractiveness to the community.

## 5.2 Paper Disruptiveness

The significant growth of MOEA/D literature not necessarily equates to increased innovative knowledge, as many papers may simply echo or refine established ideas rather than introducing disruptive paradigms. To discern the disruptive nature of MOEA/D publications, we apply the CD index, a metric devised by the scientometrics community [102]. This index evaluates a paper's influence by examining how its citations relate to those of its predecessors. A paper with a CD index approaching 1 is considered disruptive, indicating that subsequent work citing it tends not to cite its antecedents, thereby marking a deviation from prior knowledge. On the other hand, a CD index near $-1$ signifies a consolidating paper, which suggests later citations also acknowledge its precursors, thus reinforcing existing research frameworks. Employing the CD index enables an assessment of whether MOEA/D literature predominantly consolidates existing knowledge or serves as a catalyst for research innovation.

From the results shown on the left of Fig. 15, we find that MOEA/D research is becoming less disruptive over time, with the CD index dropping from 0.54 in 2009 to 0.10 in 2022. This observation also resonates with linguistic analysis, wherein the introduction of new words in paper abstracts and titles is becoming less frequent, as indicated by the declining type-token ratio [103], measured by unique/total words shown in Fig. 15 (Right). These observations are consistent with general trends in the global scientific literature, and can be partially attributed to a dearth of 'low-hanging fruit' [103, 104].

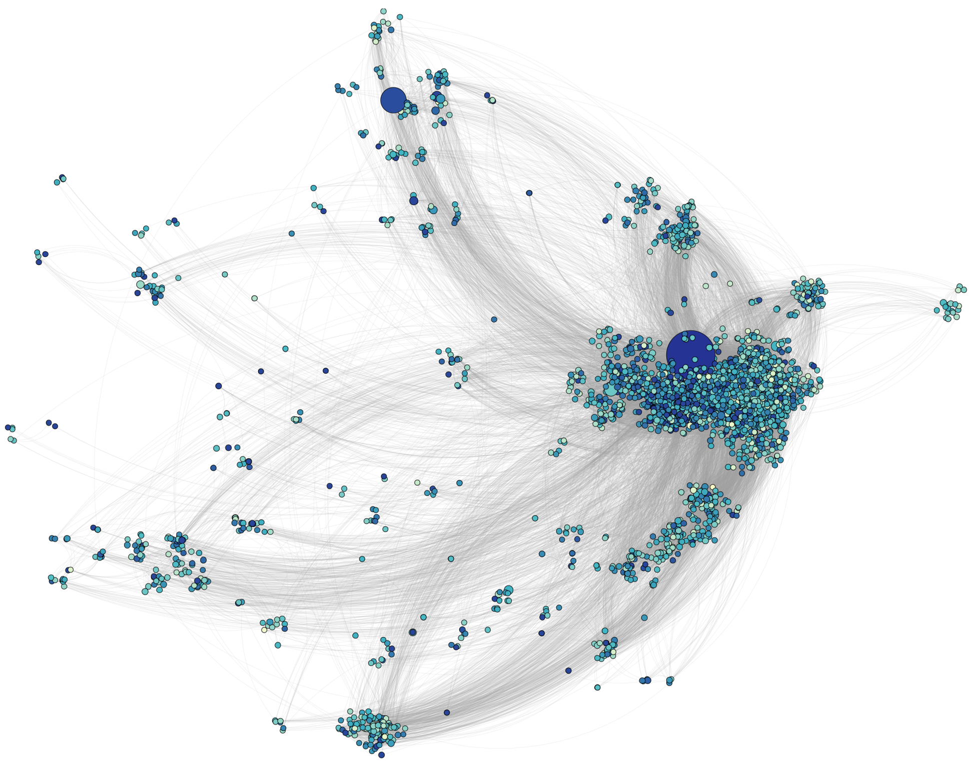

*Figure 16.* MOEA/D citation network visualized via a force-directed layout. Node size and color corresponds to the PageRank centrality of each paper.

### 5.3 Community Structure

In addition to the evolutionary dynamics, MOEA/D citation network also reveals a highly modular structure (see Fig. 16), i.e., papers are more densely connected to each other within each community than those outside it. Such communities are often formulated based on three types of citation relationships:

- Direct citation: It is the most direct relationship between papers, where the citing paper may have utilized, built upon, or discussed the content of the cited one. It is the fundamental building block of the evolution of science.
- Co-citation [105]: It occurs when two papers are simultaneously cited in one or more subsequent papers, which indicates potential overlaps in research content or focus.
- Bibliographic coupling [106]: In contrast to co-citation, bibliographic coupling examines historical connections where two papers share common references, which establishes a similar background or context for two works.

Exploring citation communities within the MOEA/D domain offers a unique perspective, potentially validating or complementing insights from our previous topic modeling efforts. To this end, we first applied a community detection algorithm to examine the underlying structure of these communities. We then develop a compact visualization of the network with detected communities.

To identify communities within the citation network, we utilized the well-known `Louvain` method [107], renowned for its iterative approach to optimizing modularity [108] by aggregating closely-knit communities. The parameter `resolution` therein allows for adjusting the granularity of detected communities, enabling tailored analysis of the network's structure.

When setting `resolution` as 3, the `Louvain` method identified 69 communities, comparable to the number of topics obtained in the topic modeling. The results, depicted in Fig. 17, use a compressed network visualization where each node corresponds to an identified community. These research themes generally mirror those discovered through `BERTopic`. Furthermore, a detailed comparison of individual

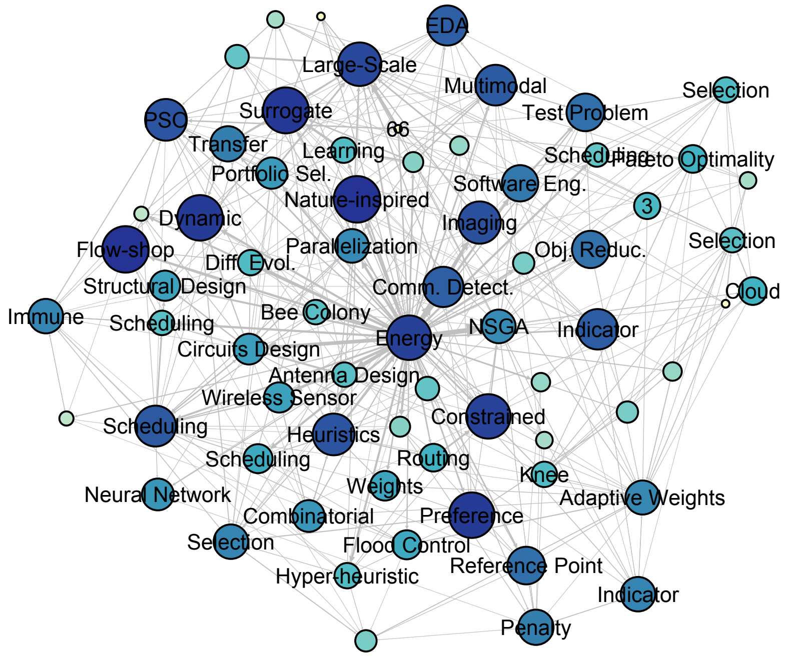


*Figure 17.* Compressed representation of MOEA/D citation network. Each node is a detected community by `Louvain` algorithm with `resolution=3`, and node size & color corresponds to number of papers in each community. Edges represent inter-community citations.

topics reveals substantial overlaps between papers grouped by both `Louvain` and `BERTopic`. For instance, topics on preference, reference vectors, and knee points demonstrate Jaccard similarities of 0.77, 0.65, and 0.57, respectively, underscoring a high level of agreement between the two methods.

However, `Louvain` also outputs some communities that are seemingly very different from `BERTopic`. For example, the largest community detected here is labeled as 'nature-inspired algorithms', which not only covers topics like whale and wolf optimizers that are identified as independent topics previously, but also incorporates other other ones such as salp swarm algorithm [109] and bat algorithm [110]. Another example is the communities regarding heuristics, combinatorial optimization, and hyper-heuristics. Papers in these groups are often assigned to one of the application topics by `BERTopic`. Also, some other topics mentioned in Part I, such as differential evolution, are also detected as separate communities by `Louvain`.

Such difference between the results from `BERTopic` and `Louvain` primarily stems from their distinct inherent mechanisms for evaluating paper similarities.

- Topic modeling in `BERTopic` leverages sentence embedding models to investigate semantic resemblances in paper content, which are generally trained on broad text corpora. This training enables them to excel at identifying overarching themes, particularly application domains, by capturing semantic similarities but might overlook nuanced methodological distinctions that demand deep expertise.
- In contrast, community detection algorithms focus on the structural connections among papers, and are good at identifying research streams or groups. These groups, though possibly divergent in content, often converge on similar methodological concerns. This structural approach, however, may sacrifice the *coherence* of communities compared to the semantic richness captured by topic modeling.

Albeit these differences, the results yielded by the two methods are largely consistent. By additionally considering the new topics identified here, we now have a more comprehensive view of the MOEA/D research landscape.

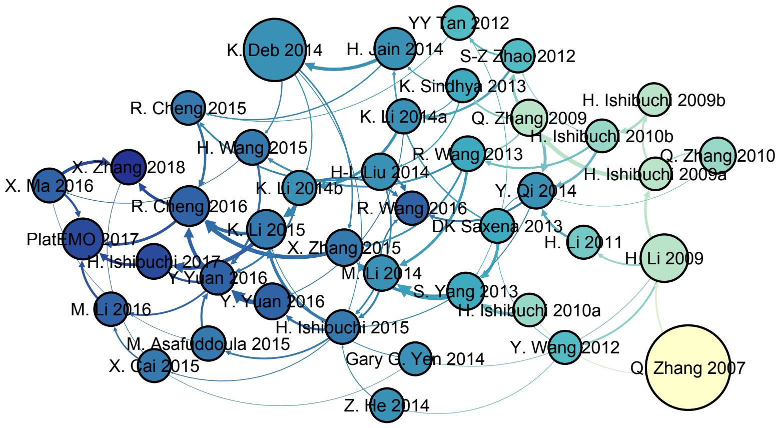


*Figure 18.* The most essential 40 nodes in the MOEA/D citation network. Node size is proportional to number of citations, color indicates publication year. Edge width implies paper similarity based on co-citation and bibliographic coupling. Paper information can be found in APPENDIX C.

### 5.4 Main Path Analysis

In addition to identifying communities within citation networks, another interesting line of work in bibliometric analysis is to delineate the evolution of a field through the extraction of a *main path* from the network. This main path aims to spotlight landmark studies and their interconnections [111–113]. However, given the intricate citation networks characteristic of domains such as MOEA/D, a singular path—or even variations like citation trees [21]—may not adequately represent the multifaceted relationships among papers. While subgraphs featuring the most influential papers could potentially depict more complex knowledge flows, their interpretability tends to diminish with an increase in the number of edges.

To address these challenges, we propose a novel approach for main path analysis, which consists of three key steps.

**Identifying essential works.** We first employ the ranking method proposed in [114] to identify a group of the most influential papers in the citation network. It simultaneously considers: ▶ paper citations relationships (i.e., number of citations, and whether the citing papers themselves are influential); ▶ author authority; and ▶ venue prestige during the ranking by leveraging the knowledge graph described in Section 2. In addition, the dynamic nature of the network is also taken into account to alleviate the bias to earlier, established works. It can thus provide more reasonable ranking results compared to other ranking methods like Cite-Rank [115] and P-Rank [116].

**Network trimming.** We then extract a subgraph containing the most essential papers. We further trim this subgraph by allowing papers that cite others to also 'inherit' their references, thereby eliminating redundant connections to earlier works such as MOEA/D when they are already acknowledged by the cited papers. By doing so, the network can better focus on paths that lead to new knowledge.

**Edge weighting.** To further enhance interpretability, we adopt the concept of co-citation and bibliographic coupling introduced in Section 5.3 to assign weights to the edges in the obtained subgraph. For each pair of papers, we count their co-citations and the Jaccard similarity of their set of references. We then use the normalized sum of these two measures as the weight of the path between the two papers, where a larger weight indicates higher relevance and similarity.

Fig. 18 illustrates the 'backbone' of the MOEA/D citation network extracted by our method, which mainly comprises papers published in the 10 years after the inception of MOEA/D that have established their influences. Over half of these works have been reviewed in PART I of this survey (see APPENDIX C for detailed information). Overall we see that most works are built upon two foundational works: MOEA/D in 2007 [1], and NSGA-III in 2014 [57, 117]. The presented studies do not exhibit typical sequential patterns, but rather form a complex interwoven structure. A primary reason here is that the MOEA/D literature is often closely related to the broader EMO domain, and thereby studies in both areas are frequently connected. Notably, works by *Prof.* Hisao Ishibuchi, e.g., [118, 119], often

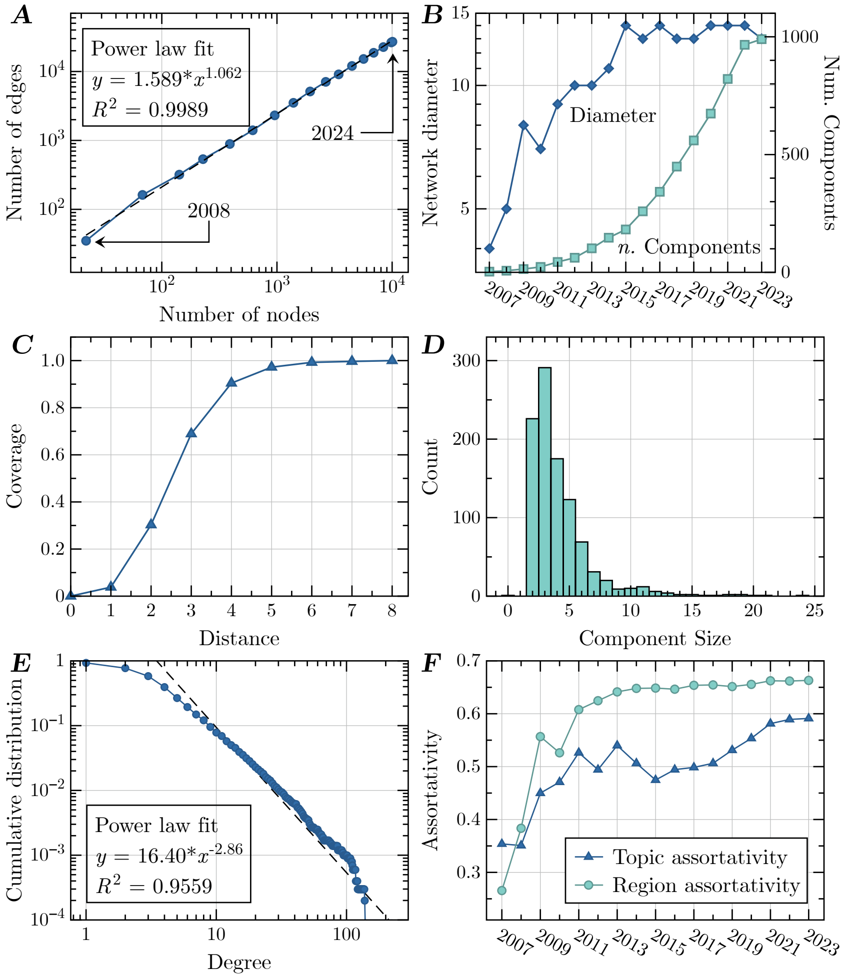


*Figure 19.* General pattterns for MOEA/D author collaboration networks. **(A)** Network growth during 2008 and 2024. **(B)** Network diameter (largest connected component) and number of connected components each year. **(C)** Percentages of nodes that can be reached within different numbers of hubs from the highest-degree node in the largest connected component. **(D)** Distribution of the size of the connected components in the network, excluding the largest one. **(E)** Cumulative distribution of author connections within the network. **(F)** Assortativity coefficient for author nationality and primary research interest.

serve as bridges between different research streams. These works are usually based on well-motivated purpose-built experiments that offer insights into the behavior of EMO algorithms and spark further research.

# 6 Collaboration Network Analysis

In this section, we analyze the author collaboration network, another important component of the MOEA/D knowledge graph. Similar to the previous sections, we begin by examining various network metrics to illuminate the composition and evolution of the MOEA/D research community. Thereafter, we spotlight the most active researchers in this domain and dissect their collaboration patterns.

## 6.1 Network Patterns

### 6.1.1 Network components

Fig. 19(B) also shows the evolving number of connected components within the network. We find that the MOEA/D community is fragmented into a significant number of isolated research groups—approximately 1,000 at the end of 2023. These groups are predominantly small, with most comprising $< 5$ researchers shown in Fig. 19(D), and they mainly focus on application-oriented research. The largest of these isolated groups includes 24 researchers. In contrast, the network's main component encompasses more than $5,700$ authors. This indicates a vast difference in scale and possibly in research focus within the community.

#### 6.1.2 Network growth and small world effect

As shown in Fig. 19(A), the MOEA/D community has been growing at an accelerated pace in recent years, with a power-law exponent $\alpha = 1.06$. This indicates that, as the network expands, the density of connections between researchers intensifies, prompting the community to exhibit characteristics of 'small-world' networks [120]. A tangible example of this phenomenon is the stable network diameter $d$ shown in Fig. 19(B). Despite the network's continual expansion, the diameter has remained constant at approximately 14 since 2015, which is relatively short considering the scale of the network. Moreover, as shown in Fig. 19(C), in the largest component, around 70% of authors can be reached within just 3 hops from the highest-degree node, and nearly all authors are within 5 hops. This compact connectivity suggests that the MOEA/D research community is becoming increasingly cohesive, facilitating easier access to collaborators and knowledge dissemination across the network.

#### 6.1.3 Degree distribution

As depicted in Fig. 19(E), the degree distribution in the author collaboration network exhibits a highly skewed pattern, similar to what is observed in citation networks. The majority of authors are connected to only a small number of other researchers. These typically include students, who are most frequently connected to their supervisors and immediate colleagues. This pattern, characterized by numerous low-degree nodes, forms the foundational structure of the entire network, suggesting a significant presence of tightly-knit research groups. Note that it is important to understand this distribution, because it reflects on the hierarchical and potentially insular nature of collaborations within the community, while also highlighting the pivotal role of more connected individuals in bridging disparate groups.

#### 6.1.4 Mixing patterns

Another dimension of our investigation is to understand whether MOEA/D researchers prefer collaborating with others in their geographic region or those working on similar topics. Analysis of the assortativity coefficients [121] in Fig. 19(F) shows an increasing trend over time for both geographical and topical collaboration preferences. This suggests a growing inclination within research groups towards more conservative collaboration patterns and focused areas of interest. While this may foster in-depth exploration within established research domains, it is expected to promote broader interdisciplinary and cross-regional partnerships, which have been shown to infuse fresh perspectives and foster innovation in a research community [122, 123].

### 6.2 Active researchers

To identify the most influential authors within the MOEA/D community, we utilize Google's PageRank centrality [124], a metric widely acknowledged for its effectiveness in highlighting key figures in networked environments such as citation networks [125]. To ensure a robust comparison, we complement this with additional centrality measures as follows.

- Degree centrality: It reflects the *quantity* of an author's connections within the community. Authors with high-degree centrality are considered central figures. However, this may not fully reflect the influence of an author since it overlooks the connection's *strength* and *quality*.
- Weighted degree centrality: By factoring in the *strength* of collaborations, this measure offers a nuanced measure an authors' connections in the network. Yet, it still does not account for the qualitative aspect of connections.

*Table 2.* 50 Most active authors in the MOEA/D research community.

| Author Name | PageRank | Degree | w_Degree | #Citations | #Papers | Betweenness | YoE |
|---|---|---|---|---|---|---|---|
| Zhang, Qingfu | 3.5E−3 | 227 | 111.5 | 12,329 | 150 | 5.9E−2 | 2007 |
| Jin, Yaochu | 2.0E−3 | 111 | 70.4 | 9,019 | 94 | 2.1E−2 | 2008 |
| Deb, Kalyanmoy | 2.0E−3 | 108 | 53.0 | 9,286 | 78 | 1.3E−2 | 2009 |
| Ishibuchi, Hisao | 2.0E−3 | 102 | 106.4 | 4,249 | 150 | 1.3E−2 | 2009 |
| Yang, Shengxiang | 1.8E−3 | 118 | 59.2 | 3,715 | 78 | 1.7E−2 | 2013 |
| Gong, Maoguo | 1.8E−3 | 112 | 57.6 | 3,163 | 74 | 1.2E−2 | 2009 |
| Tan, Kay Chen | 1.7E−3 | 139 | 54.5 | 3,146 | 72 | 2.8E−2 | 2011 |
| Coello Coello, C. A. | 1.7E−3 | 118 | 53.9 | 3,192 | 83 | 3.4E−2 | 2010 |
| Jiao, Licheng | 1.7E−3 | 122 | 50.5 | 3,242 | 64 | 3.4E−2 | 2009 |
| Zhang, Jun | 1.6E−3 | 138 | 44.8 | 3,080 | 57 | 2.6E−2 | 2011 |
| Wang, Rui | 1.4E−3 | 93 | 41.5 | 2,538 | 55 | 1.5E−2 | 2012 |
| Yen, Gary G. | 1.4E−3 | 76 | 38.2 | 3,036 | 56 | 1.1E−2 | 2011 |
| Wang, Ling | 1.4E−3 | 95 | 38.1 | 1,705 | 52 | 1.4E−2 | 2017 |
| Lin, Qiuzhen | 1.4E−3 | 78 | 53.2 | 1,767 | 66 | 2.0E−3 | 2015 |
| Zhang, Xingyi | 1.2E−3 | 67 | 42.4 | 4,698 | 54 | 4.5E−3 | 2015 |
| Yao, Xin | 1.1E−3 | 86 | 33.6 | 3,821 | 45 | 1.6E−2 | 2011 |
| Gao, Liang | 1.1E−3 | 67 | 31.7 | 894 | 41 | 5.9E−3 | 2016 |
| Li, Miqing | 1.1E−3 | 74 | 34.2 | 3,284 | 46 | 8.1E−3 | 2013 |
| Gong, Dunwei | 1.1E−3 | 86 | 29.7 | 1,477 | 38 | 1.3E−2 | 2014 |
| Zheng, Jinhua | 1.0E−3 | 61 | 36.5 | 1,297 | 46 | 3.1E−3 | 2013 |
| Zhou, Aimin | 1.0E−3 | 61 | 32.1 | 3,119 | 45 | 8.5E−3 | 2008 |
| Zou, Juan | 1.0E−3 | 65 | 35.3 | 477 | 44 | 3.5E−3 | 2017 |
| Tian, Ye | 9.9E−4 | 57 | 36.0 | 4,540 | 46 | 7.8E−3 | 2015 |
| Liu, Hai-Lin | 9.5E−4 | 38 | 31.3 | 1,205 | 46 | 1.5E−3 | 2012 |
| Nojima, Yusuke | 9.2E−4 | 35 | 52.3 | 2,381 | 72 | 2.9E−4 | 2009 |
| Li, Ke | 8.9E−4 | 53 | 26.7 | 2,523 | 37 | 5.8E−3 | 2013 |
| Liu, Jing | 8.6E−4 | 53 | 22.9 | 889 | 34 | 5.4E−3 | 2013 |
| Cai, Xinye | 8.5E−4 | 62 | 27.4 | 1,282 | 35 | 6.5E−3 | 2013 |
| Zhang, Tao | 8.3E−4 | 46 | 23.2 | 1,223 | 31 | 5.2E−3 | 2013 |
| Mirjalili, Seyedali | 8.0E−4 | 58 | 16.0 | 5,040 | 21 | 7.6E−3 | 2015 |
| Gong, Wenyin | 8.0E−4 | 38 | 23.8 | 369 | 32 | 1.6E−3 | 2015 |
| Wang, Lei | 7.9E−4 | 52 | 20.1 | 366 | 26 | 7.3E−3 | 2013 |
| Das, Swagatam | 7.7E−4 | 52 | 18.1 | 2,083 | 25 | 6.7E−3 | 2011 |
| Zhang, Mengjie | 7.7E−4 | 40 | 22.3 | 626 | 31 | 2.8E−3 | 2017 |
| Wang, Yuping | 7.5E−4 | 34 | 21.7 | 459 | 32 | 3.7E−3 | 2008 |
| Liu, Ruochen | 7.2E−4 | 37 | 20.2 | 411 | 27 | 1.1E−3 | 2014 |
| Cheng, Ran | 7.1E−4 | 42 | 26.9 | 5,017 | 35 | 1.0E−3 | 2015 |
| Zhang, Kai | 7.0E−4 | 52 | 18.4 | 558 | 24 | 3.7E−3 | 2015 |
| Cao, Bin | 7.0E−4 | 44 | 18.4 | 1,587 | 23 | 2.2E−3 | 2017 |
| Zhou, Mengchu | 6.7E−4 | 54 | 17.2 | 1,090 | 4 | 6.0E−3 | 2017 |
| Li, Hui | 6.7E−4 | 45 | 21.7 | 4,719 | 30 | 1.1E−2 | 2008 |
| Kwong, Sam | 6.7E−4 | 47 | 19.4 | 2,150 | 25 | 4.9E−3 | 2012 |
| Ong, Yew-Soon | 6.7E−4 | 49 | 19.7 | 1,983 | 26 | 4.0E−3 | 2011 |
| Fu, Yaping | 6.6E−4 | 33 | 17.4 | 635 | 23 | 6.3E−4 | 2014 |
| Fan, Zhun | 6.5E−4 | 46 | 21.4 | 1,093 | 27 | 2.9E−3 | 2014 |
| Zhu, Zexuan | 6.2E−4 | 63 | 19.3 | 785 | 24 | 9.4E−3 | 2015 |
| Wang, Jiahai | 6.2E−4 | 42 | 17.2 | 912 | 23 | 1.8E−3 | 2013 |
| Zhou, Yuren | 6.2E−4 | 29 | 22.1 | 934 | 31 | 1.7E−3 | 2017 |
| Dai, Cai | 6.2E−4 | 24 | 16.7 | 322 | 28 | 1.3E−3 | 2014 |
| Wang, Handing | 6.1E−4 | 42 | 17.9 | 1,579 | 27 | 5.0E−3 | 2013 |

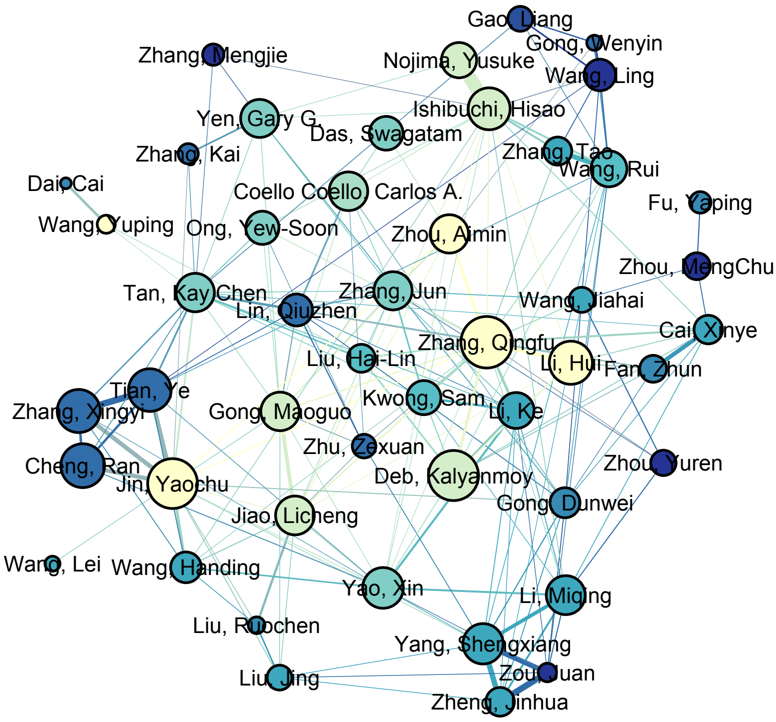


*Figure 20.* Local community of top-50 active authors in the MOEA/D research community. Node size and color are proportional to the author's PageRank centrality and year of entry, respectively. Edge width implies collaboration strength, as defined in Section 2.4.

- PageRank centrality: It adds a layer by valuing the *quality* of connections, recognizing authors linked to well-regarded peers as more influential. Therefore, it pinpoints pivotal nodes within the network.
- Betweenness centrality: It measures an author's role as a knowledge broker, identifying those who bridge diverse groups within the MOEA/D ecosystem.

These measures together enable a comprehensive assessment of author influence. They indicate not only who is central but also who facilitates knowledge flow and community cohesion in MOEA/D research. Table 2 presents the top-50 authors in the MOEA/D community, ranked according to various centrality measures and bibliometric indicators. Notably, researchers with the highest PageRank centrality are Qingfu Zhang, Yaochu Jin, Kalyanmoy Deb, Hisao Ishibuchi, and Shengxiang Yang. These authors typically also ranked high by other metrics, and thereby play a crucial role in the community. In particular, *Prof.* Qingfu Zhang, one of the co-authors of the seminal paper of MOEA/D, dominates the rankings in all measures assessed.

We also visualize the collaboration network of these top-50 authors in Fig. 20. From this plot, we find that they are closely connected to each other. Notably, within this network, cohesive local groups are prevalent, with 26 4-cliques and 14 5-cliques. Notably, authors who serve as bridges between these groups, such as Carlos A. Coello Coello and Kay Chen Tan, often have high betweenness centrality. Central figures in the network, especially Qingfu Zhang, have strong collaborative ties with Hui Li, Aimin Zhou, Kalyanmoy Deb, Ke Li, and Sam Kwong. Within the whole network, the strongest tie is between *Prof.* Hisao Ishibuchi and *Prof.* Yusuke Nojima.

**Remark 6**

As we only considered the collaboration network for this ranking, the results are only indicative of the authors' social connections in the community, and should not be considered as a measure of research impact or academic reputation.

## Acknowledgment

This work was supported in part by the UKRI Future Leaders Fellowship under Grant MR/S017062/1 and MR/X011135/1; in part by NSFC under Grant 62376056 and 62076056; in part by the Royal Society under Grant IES/R2/212077; in part by the EPSRC under Grant 2404317; in part by the Kan Tong Po Fellowship (KTP\R1\231017); and in part by the Amazon Research Award and Alan Turing Fellowship.

## References

[1] Q. Zhang and H. Li, "MOEA/D: A multiobjective evolutionary algorithm based on decomposition," *IEEE Trans. Evol. Comput.*, vol. 11, no. 6, pp. 712–731, 2007. 1, 3, 9, 23

[2] D. Budgen and P. Brereton, "Performing systematic literature reviews in software engineering," in *ICSE'06: Proc, of the 28th International Conference on Software Engineering.* ACM, 2006, pp. 1051–1052. 1

[3] D. Moher, L. Shamseer, M. Clarke, D. Ghersi, A. Liberati, M. Petticrew, P. Shekelle, L. A. Stewart, and P.-P. Group, "Preferred reporting items for systematic review and meta-analysis protocols (prisma-p) 2015 statement," *Syst. Rev.*, vol. 4, no. 1, p. 1, 2015. 2

[4] R. Churchill and L. Singh, "The evolution of topic modeling," *ACM Comput. Surv.*, vol. 54, no. 10s, nov 2022. 2, 3, 9

[5] M. Falkenberg, A. Galeazzi, M. Torricelli, N. Di Marco, F. Larosa, M. Sas, A. Mekacher, W. Pearce, F. Zollo, W. Quattrociocchi, and A. Baronchelli, "Growing polarization around climate change on social media," *Nature Climate Change*, vol. 12, no. 12, pp. 1114–1121, 2022. 2

[6] A. Ameur, S. Hamdi, and S. B. Yahia, "Sentiment analysis for hotel reviews: A systematic literature review," *ACM Comput. Surv.*, vol. 56, no. 2, pp. 51:1–51:38, 2024. 2

[7] S. Fortunato and D. Hric, "Community detection in networks: A user guide," *Phy. Rep.*, vol. 659, pp. 1–44, 2016. 2

[8] P. Bedi and C. Sharma, "Community detection in social networks," *WIREs Data Mining Knowl. Discov.*, vol. 6, no. 3, pp. 115–135, 2016. 2

[9] C. A. Brittin, S. J. Cook, D. H. Hall, S. W. Emmons, and N. Cohen, "A multi-scale brain map derived from whole-brain volumetric reconstructions," *Nature*, vol. 591, no. 7848, pp. 105–110, 2021. 2

[10] S. Minaee, T. Mikolov, N. Nikzad, M. Chenaghlu, R. Socher, X. Amatriain, and J. Gao, "Large language models: A survey," *CoRR*, vol. abs/2402.06196, 2024. 2

[11] Z. Tan, A. Beigi, S. Wang, R. Guo, A. Bhattacharjee, B. Jiang, M. Karami, J. Li, L. Cheng, and H. Liu, "Large language models for data annotation: A survey," *CoRR*, vol. abs/2402.13446, 2024. 2

[12] J. S. Huh, C. Shin, and E. Choi, "Pool-search-demonstrate: Improving data-wrangling LLMs via better in-context examples," in *TPL'23: Proc. of the NeurIPS 2023 Second Table Representation Learning Workshop*, 2023. [Online]. Available: https://openreview.net/forum?id=6Kb3pE9nWQ 2

[13] Y. Chang, K. Lo, T. Goyal, and M. Iyyer, "Booookscore: A systematic exploration of book-length summarization in the era of LLMs," in *ICLR'24: Proc. of the 12th International Conference on Learning Representations*, 2024. 2

[14] J. S. Liu, L. Y. Lu, W.-M. Lu, and B. J. Lin, “Data envelopment analysis 1978–2010: A citation-based literature survey,” *Omega*, vol. 41, no. 1, pp. 3–15, 2013. 2, 3

[15] S. Heldens, P. Hijma, B. van Werkhoven, J. Maassen, A. S. Z. Belloum, and R. van Nieuwpoort, “The landscape of exascale research: A data-driven literature analysis,” *ACM Comput. Surv.*, vol. 53, no. 2, pp. 23:1–23:43, 2021. 2, 3, 10

[16] L. Sun and Y. Yin, “Discovering themes and trends in transportation research using topic modeling,” *Transportation Research Part C: Emerging Technologies*, vol. 77, pp. 49–66, 2017. 2, 3, 10

[17] J. J. Jeong, Y. Zolotavkin, and R. Doss, “Examining the current status and emerging trends in continuous authentication technologies through citation network analysis,” *ACM Comput. Surv.*, vol. 55, no. 6, pp. 122:1–122:31, 2023. 2, 3

[18] S. Fortunato, C. T. Bergstrom, K. Börner, J. A. Evans, D. Helbing, S. Milojević, A. M. Petersen, F. Radicchi, R. Sinatra, B. Uzzi, A. Vespignani, L. Waltman, D. Wang, and A.-L. Barabási, “Science of science,” *Science*, vol. 359, no. 6379, p. eaao0185, 2018. 2, 3, 19

[19] M. B. Ross, B. M. Glennon, R. Murciano-Goroff, E. G. Berkes, B. A. Weinberg, and J. I. Lane, “Women are credited less in science than men,” *Nature*, vol. 608, no. 7921, pp. 135–145, 2022. 2, 3

[20] R. Sinatra, D. Wang, P. Deville, C. Song, and A.-L. Barabási, “Quantifying the evolution of individual scientific impact,” *Science*, vol. 354, no. 6312, p. aaf5239, 2016. 2, 3

[21] L. Wu, D. Wang, and J. A. Evans, “Large teams develop and small teams disrupt science and technology,” *Nature*, vol. 566, no. 7744, pp. 378–382, 2019. 2, 3, 23

[22] W. Li, T. Aste, F. Caccioli, and G. Livan, “Early coauthorship with top scientists predicts success in academic careers,” *Nature Communications*, vol. 10, no. 1, p. 5170, 2019. 2, 3

[23] C. Ihle, D. Trautwein, M. Schubotz, N. Meuschke, and B. Gipp, “Incentive mechanisms in peer-to-peer networks - A systematic literature review,” *ACM Comput. Surv.*, vol. 55, no. 14s, pp. 308:1–308:69, 2023. 3

[24] H. Korala, D. Georgakopoulos, P. P. Jayaraman, and A. Yavari, “A survey of techniques for fulfilling the time-bound requirements of time-sensitive iot applications,” *ACM Comput. Surv.*, vol. 54, no. 11s, pp. 228:1–228:36, 2022. 3

[25] J. Klaiber and C. van Dinther, “Deep learning for variable renewable energy: A systematic review,” *ACM Comput. Surv.*, vol. 56, no. 1, pp. 1:1–1:37, 2024. 3

[26] K. Lo, L. L. Wang, M. Neumann, R. Kinney, and D. S. Weld, “S2ORC: the semantic scholar open research corpus,” in *ACL'20: Proc. of the 58th Annual Meeting of the Association for Computational Linguistics.* ACL, 2020, pp. 4969–4983. 5

[27] A. Perianes-Rodríguez, L. Waltman, and N. J. van Eck, “Constructing bibliometric networks: A comparison between full and fractional counting,” *J. Informetrics*, vol. 10, no. 4, pp. 1178–1195, 2016. 6

[28] A. Cohan, W. Ammar, M. van Zuylen, and F. Cady, “Structural scaffolds for citation intent classification in scientific publications,” in *NAACL-HLT'19: Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies.* ACL, 2019, pp. 3586–3596. 9

[29] M. Roman, A. Shahid, S. Khan, A. Koubaa, and L. Yu, “Citation intent classification using word embedding,” *IEEE Access*, vol. 9, pp. 9982–9995, 2021. 9

[30] S. N. Kunnath, D. Pride, and P. Knoth, “Prompting strategies for citation classification,” in *CIKM’23: Proc. of the 32nd ACM International Conference on Information and Knowledge Management*. ACM, 2023, pp. 1127–1137. 9

[31] A. Sidorova, N. Evangelopoulos, J. S. Valacich, and T. Ramakrishnan, “Uncovering the intellectual core of the information systems discipline,” *MIS Quarterly*, vol. 32, no. 3, pp. 467–482, 2008. 9

[32] G. Mathew, A. Agrawal, and T. Menzies, “Finding trends in software research,” *IEEE Trans. Software Eng.*, vol. 49, no. 4, pp. 1397–1410, 2023. 10

[33] M. J. Paul and M. Dredze, “Discovering health topics in social media using topic models,” *PLOS ONE*, vol. 9, no. 8, pp. 1–11, 08 2014. 10

[34] F. Shahnaz, M. W. Berry, V. Pauca, and R. J. Plemmons, “Document clustering using nonnegative matrix factorization,” *Inf. Process. Manage.*, vol. 42, no. 2, pp. 373–386, 2006. 10

[35] D. M. Blei, A. Y. Ng, and M. I. Jordan, “Latent dirichlet allocation,” *J. Mach. Learn. Res.*, vol. 3, pp. 993–1022, 2003. 10

[36] A. Abdelrazek, Y. Eid, E. Gawish, W. Medhat, and A. H. Yousef, “Topic modeling algorithms and applications: A survey,” *Inf. Syst.*, vol. 112, p. 102131, 2023. 10

[37] J. Devlin, M. Chang, K. Lee, and K. Toutanova, “BERT: pre-training of deep bidirectional transformers for language understanding,” in *NAACL-HLT’19: Proc. of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. ACM, 2019, pp. 4171–4186. 10

[38] J. Lee, W. Yoon, S. Kim, D. Kim, S. Kim, C. H. So, and J. Kang, “Biobert: a pre-trained biomedical language representation model for biomedical text mining,” *Bioinform.*, vol. 36, no. 4, pp. 1234–1240, 2020. 10

[39] Y. Liu, M. Ott, N. Goyal, J. Du, M. Joshi, D. Chen, O. Levy, M. Lewis, L. Zettlemoyer, and V. Stoyanov, “Roberta: A robustly optimized BERT pretraining approach,” *CoRR*, vol. abs/1907.11692, 2019. 10

[40] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, and R. Soricut, “ALBERT: A lite BERT for self-supervised learning of language representations,” in *ICLR’20: Proc. of the 8th International Conference on Learning Representations*. OpenReview.net, 2020. 10

[41] M. Grootendorst, “Bertopic: Neural topic modeling with a class-based TF-IDF procedure,” *CoRR*, vol. abs/2203.05794, 2022. 10

[42] S. Sia, A. Dalmia, and S. J. Mielke, “Tired of topic models? clusters of pretrained word embeddings make for fast and good topics too!” in *EMNLP’20: Proc. of the 2020 Conference on Empirical Methods in Natural Language Processing*. ACL, 2020, pp. 1728–1736. 10

[43] D. Angelov, “Top2vec: Distributed representations of topics,” *CoRR*, vol. abs/2008.09470, 2020. 10

[44] H. Su, W. Shi, J. Kasai, Y. Wang, Y. Hu, M. Ostendorf, W. Yih, N. A. Smith, L. Zettlemoyer, and T. Yu, “One embedder, any task: Instruction-finetuned text embeddings,” in *ACL’23: Findings of the Association for Computational Linguistics*. ACL, 2023, pp. 1102–1121. 10

[45] K. Li, K. Deb, Q. Zhang, and S. Kwong, “An evolutionary many-objective optimization algorithm based on dominance and decomposition,” *IEEE Trans. Evol. Comput.*, vol. 19, no. 5, pp. 694–716, 2015. 10

[46] N. Reimers and I. Gurevych, "Sentence-bert: Sentence embeddings using siamese bert-networks," in *EMNLP-IJCNLP'19: Proc. of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing.* ACL, 2019, pp. 3980–3990. 10

[47] N. Muennighoff, N. Tazi, L. Magne, and N. Reimers, "MTEB: massive text embedding benchmark," in *EACL'23: Proc. of the 17th Conference of the European Chapter of the Association for Computational Linguistics.* ACL, 2023, pp. 2006–2029. 10

[48] C. C. Aggarwal, A. Hinneburg, and D. A. Keim, "On the surprising behavior of distance metrics in high dimensional spaces," in *ICDT'01: Proc. of the 8th International Conference*, ser. Lecture Notes in Computer Science, vol. 1973. Springer, 2001, pp. 420–434. 11

[49] L. McInnes, J. Healy, N. Saul, and L. Großberger, "UMAP: uniform manifold approximation and projection," *J. Open Source Softw.*, vol. 3, no. 29, p. 861, 2018. 11

[50] A. Draganov, J. R. Jørgensen, K. Scheel, D. Mottin, I. Assent, T. Berry, and Ç. Aslay, "Actup: Analyzing and consolidating tsne and UMAP," in *IJCAI'23: Proc. of the Thirty-Second International Joint Conference on Artificial Intelligence.* ijcai.org, 2023, pp. 3651–3658. 11

[51] L. McInnes, J. Healy, and S. Astels, "hdbscan: Hierarchical density based clustering," *J. Open Source Softw.*, vol. 2, no. 11, p. 205, 2017. 11

[52] T. Joachims, "A probabilistic analysis of the rocchio algorithm with TFIDF for text categorization," in *ICML'97: Proc. of the Fourteenth International Conference on Machine Learning*, D. H. Fisher, Ed. Morgan Kaufmann, 1997, pp. 143–151. 11

[53] K. Li, Q. Zhang, S. Kwong, M. Li, and R. Wang, "Stable matching-based selection in evolutionary multiobjective optimization," *IEEE Trans. Evol. Comput.*, vol. 18, no. 6, pp. 909–923, 2014. 12

[54] M. Asafuddoula, T. Ray, and R. A. Sarker, "A decomposition-based evolutionary algorithm for many objective optimization," *IEEE Trans. Evol. Comput.*, vol. 19, no. 3, pp. 445–460, 2015. 12

[55] Y. Yuan, H. Xu, B. Wang, B. Zhang, and X. Yao, "Balancing convergence and diversity in decomposition-based many-objective optimizers," *IEEE Trans. Evol. Comput.*, vol. 20, no. 2, pp. 180–198, 2016. 12

[56] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan, "A fast and elitist multiobjective genetic algorithm: NSGA-II," *IEEE Trans. Evol. Comput.*, vol. 6, no. 2, pp. 182–197, 2002. 12

[57] K. Deb and H. Jain, "An evolutionary many-objective optimization algorithm using reference-point-based nondominated sorting approach, Part I: solving problems with box constraints," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 577–601, 2014. 12, 23

[58] J. Blank and K. Deb, "Pymoo: Multi-objective optimization in python," *IEEE Access*, vol. 8, pp. 89 497–89 509, 2020. 13

[59] Y. Tian, R. Cheng, X. Zhang, and Y. Jin, "Platemo: A MATLAB platform for evolutionary multi-objective optimization [educational forum]," *IEEE Comput. Intell. Mag.*, vol. 12, no. 4, pp. 73–87, 2017. 13

[60] W. Peng and Q. Zhang, "A decomposition-based multi-objective particle swarm optimization algorithm for continuous optimization problems," in *GrC'08: Proc. of the 2008 IEEE International Conference on Granular Computing.* Hangzhou, China: IEEE, 2008, pp. 534–537. 13

[61] N. A. Moubayed, A. Petrovski, and J. A. W. McCall, "A novel smart multi-objective particle swarm optimisation using decomposition," in *PPSN XI: Proc. of the 11th International Conference on Parallel Problem Solving from Nature*, ser. Lecture Notes in Computer Science, vol. 6239. Kraków, Poland: Springer, 2010, pp. 1–10. 13

[62] S. Z. Martínez and C. A. C. Coello, "A multi-objective particle swarm optimizer based on decomposition," in *GECCO'13: Proc. of the 13th Annual Genetic and Evolutionary Computation Conference*. Dublin, Ireland: ACM, 2011, pp. 69–76. 13

[63] M. Abdel-Basset, R. Mohamed, and S. Mirjalili, "A novel whale optimization algorithm integrated with nelder-mead simplex for multi-objective optimization problems," *Knowl. Based Syst.*, vol. 212, p. 106619, 2021. 13

[64] L. Ke, Q. Zhang, and R. Battiti, "MOEA/D-ACO: A multiobjective evolutionary algorithm using decomposition and antcolony," *IEEE Trans. Cybern.*, vol. 43, no. 6, pp. 1845–1859, 2013. 13

[65] Z. Zhang, Z. Xu, S. Luan, X. Li, and Y. Sun, "Opposition-based ant colony optimization algorithm for the traveling salesman problem," *Mathematics*, vol. 8, no. 10, pp. 1–16, 2020. 13

[66] J. Ning, Q. Zhao, P. Sun, and Y. Feng, "A multi-objective decomposition-based ant colony optimisation algorithm with negative pheromone," *Journal of Experimental & Theoretical Artificial Intelligence*, 2020, accepted for publication. 13

[67] Y. Wu, Z. Qi, L. Jiang, C. Zhang, and J. Xu, "A multi-objective ant colony optimization algorithm with local optimum avoidance strategy," in *ML4CS'20: Proc. of the 3rd International Conference Machine Learning for Cyber Security*, ser. Lecture Notes in Computer Science, vol. 12487. Guangzhou, China: Springer, 2020, pp. 427–435. 13

[68] R. Shang, L. Jiao, F. Liu, and W. Ma, "A novel immune clonal algorithm for MO problems," *IEEE Trans. Evol. Comput.*, vol. 16, no. 1, pp. 35–50, 2012. 13

[69] S. Mirjalili, S. Saremi, S. M. Mirjalili, and L. dos Santos Coelho, "Multi-objective grey wolf optimizer: A novel algorithm for multi-criterion optimization," *Expert Syst. Appl.*, vol. 47, pp. 106–119, 2016. 13

[70] H. Ishibuchi, N. Akedo, and Y. Nojima, "Behavior of multiobjective evolutionary algorithms on many-objective knapsack problems," *IEEE Trans. Evol. Comput.*, vol. 19, no. 2, pp. 264–283, 2015. 14

[71] A. Zhou, F. Gao, and G. Zhang, "A decomposition based estimation of distribution algorithm for multiobjective traveling salesman problems," *Comput. Math. Appl.*, vol. 66, no. 10, pp. 1857–1868, 2013. 14

[72] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff, "Test problems for large-scale multiobjective and many-objective optimization," *IEEE Trans. Cybern.*, vol. 47, no. 12, pp. 4108–4121, 2017. 14

[73] S. M. Venske, C. P. de Almeida, R. Lüders, and M. R. Delgado, "Selection hyper-heuristics for the multi and many-objective quadratic assignment problem," *Comput. Oper. Res.*, vol. 148, p. 105961, 2022. 14

[74] S. Sengupta, S. Das, M. Nasir, A. V. Vasilakos, and W. Pedrycz, "An evolutionary multiobjective sleep-scheduling scheme for differentiated coverage in wireless sensor networks," *IEEE Trans. Syst. Man Cybern. Part C*, vol. 42, no. 6, pp. 1093–1102, 2012. 15

[75] Y. Zhou, J. Liu, Y. Zhang, and X. Gan, "A multi-objective evolutionary algorithm for multi-period dynamic emergency resource scheduling problems," *Transportation Research Part E: Logistics and Transportation Review*, vol. 99, pp. 77–95, 2017. 15

[76] P. Chang, S. Chen, Q. Zhang, and J. Lin, "MOEA/D for flowshop scheduling problems," in *CEC'08: Proc. of the IEEE Congress on Evolutionary Computation*. IEEE, 2008, pp. 1433–1438. 15

[77] S. K. Mishra, G. Panda, and R. Majhi, "A comparative performance assessment of a set of multiobjective algorithms for constrained portfolio assets selection," *Swarm Evol. Comput.*, vol. 16, pp. 38–51, 2014. 15

[78] Y. Xiang, X. Yang, Y. Zhou, and H. Huang, "Enhancing decomposition-based algorithms by estimation of distribution for constrained optimal software product selection," *IEEE Trans. Evol. Comput.*, vol. 24, no. 2, pp. 245–259, 2020. 15

[79] J. Li, J. Chen, B. Xin, and L. Dou, "Solving multi-objective multi-stage weapon target assignment problem via adaptive NSGAII and adaptive MOEA/D: A comparison study," in *CEC'15: Proc. of the IEEE Congress on Evolutionary Computation*. IEEE, 2015, pp. 3132–3139. 15

[80] Y. Qi, Z. Hou, H. Li, J. Huang, and X. Li, "A decomposition based memetic algorithm for multi-objective vehicle routing problem with time windows," *Comput. Oper. Res.*, vol. 62, pp. 61–77, 2015. 15

[81] D. Ding and G. Wang, "Modified multiobjective evolutionary algorithm based on decomposition for antenna design," *IEEE Trans. Antennas Propag.*, vol. 61, no. 10, pp. 5301–5307, 2013. 15

[82] Y. Xu, O. Ding, R. Qu, and K. Li, "Hybrid multi-objective evolutionary algorithms based on decomposition for wireless sensor network coverage optimization," *Appl. Soft Comput.*, vol. 68, pp. 268–282, 2018. 15

[83] A. Konstantinidis and K. Yang, "Multi-objective k-connected deployment and power assignment in wsns using a problem-specific constrained evolutionary algorithm based on decomposition," *Comput. Commun.*, vol. 34, no. 1, pp. 83–98, 2011. 15

[84] H. Xing, Z. Wang, T. Li, H. Li, and R. Qu, "An improved MOEA/D algorithm for multi-objective multicast routing with network coding," *Appl. Soft Comput.*, vol. 59, pp. 88–103, 2017. 15

[85] V. Ho-Huu, S. Hartjes, H. G. Visser, and R. Curran, "An improved MOEA/D algorithm for bi-objective optimization problems with complex pareto fronts and its application to structural optimization," *Expert Syst. Appl.*, vol. 92, pp. 430–446, 2018. 15

[86] P. Champasak, N. Panagant, N. Pholdee, S. Bureerat, and A. R. Yildiz, "Self-adaptive many-objective meta-heuristic based on decomposition for many-objective conceptual design of a fixed wing unmanned aerial vehicle," *Aero. Sci. Technol.*, vol. 100, p. 105783, 2020. 15

[87] S. Kumar, P. Jangir, G. G. Tejani, and P. Manoharan, "A decomposition based multi-objective heat transfer search algorithm for structure optimization," *Knowl. Based Syst.*, vol. 253, p. 109591, 2022. 15

[88] J. Jiang, F. Han, Q. Ling, J. Wang, T. Li, and H. Han, "Efficient network architecture search via multiobjective particle swarm optimization based on decomposition," *Neural Networks*, vol. 123, pp. 305–316, 2020. 16

[89] R. Jiao, B. H. Nguyen, B. Xue, and M. Zhang, "A survey on evolutionary multiobjective feature selection in classification: Approaches, applications, and challenges," *IEEE Trans. Evol. Comput.*, pp. 1–1, 2023. 16

[90] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay, "A survey of multiobjective evolutionary clustering," *ACM Comput. Surv.*, vol. 47, no. 4, pp. 61:1–61:46, 2015. 16

[91] B. Cao, J. Zhao, Z. Lv, and P. Yang, "Diversified personalized recommendation optimization based on mobile data," *IEEE Trans. Intell. Transp. Syst.*, vol. 22, no. 4, pp. 2133–2139, 2021. 16

[92] C. Liu, X. Xu, and D. Hu, "Multiobjective reinforcement learning: A comprehensive overview," *IEEE Trans. Syst. Man Cybern. Syst.*, vol. 45, no. 3, pp. 385–398, 2015. 16

[93] M. Rosen-Zvi, T. L. Griffiths, M. Steyvers, and P. Smyth, "The author-topic model for authors and documents," in *UAI'04, Proc. of the 20th Conference in Uncertainty in Artificial Intelligence.* AUAI Press, 2004, pp. 487–494. 16

[94] D. J. de Solla Price, "Networks of scientific papers," *Science*, vol. 149, no. 3683, pp. 510–515, 1965. 19

[95] T. Kuhn, M. c. v. Perc, and D. Helbing, "Inheritance patterns in citation networks reveal scientific memes," *Phys. Rev. X*, vol. 4, p. 041036, Nov 2014. 19

[96] J. Leskovec, J. Kleinberg, and C. Faloutsos, "Graphs over time: densification laws, shrinking diameters and possible explanations," in *SIGKDD'05: Proc. of the 11th ACM SIGKDD International Conference on Knowledge Continuous in Data Mining.* ACM, 2005, pp. 177–187. 19

[97] A.-L. Barabási and R. Albert, "Emergence of scaling in random networks," *Science*, vol. 286, no. 5439, pp. 509–512, 1999. 19, 20

[98] S. Redner, "How popular is your paper? an empirical study of the citation distribution," *Eur. Phys. J. B*, vol. 4, no. 2, pp. 131–134, 1998. 20

[99] D. Wang, C. Song, and A.-L. Barabási, "Quantifying long-term scientific impact," *Science*, vol. 342, no. 6154, pp. 127–132, 2013. 20

[100] L. A. N. Amaral, A. Scala, M. Barthélémy, and H. E. Stanley, "Classes of small-world networks," *Proc. Nat. Acad. Sci. USA*, vol. 97, no. 21, pp. 11 149–11 152, 2000. 20

[101] G. Bianconi and A.-L. Barabási, "Competition and multiscaling in evolving networks," *Europhys. Lett.*, vol. 54, no. 4, p. 436, 2001. 20

[102] R. J. Funk and J. Owen-Smith, "A dynamic network measure of technological change," *Manag. Sci.*, vol. 63, no. 3, pp. 791–817, 2017. 20

[103] M. Park, E. Leahey, and R. J. Funk, "Papers and patents are becoming less disruptive over time," *Nature*, vol. 613, no. 7942, pp. 138–144, 2023. 20

[104] J. S. G. Chu and J. A. Evans, "Slowed canonical progress in large fields of science," *Proceedings of the National Academy of Sciences*, vol. 118, no. 41, p. e2021636118, 2021. 20

[105] H. Small, "Co-citation in the scientific literature: A new measure of the relationship between two documents," *J. Am. Soc. Inf. Sci.*, vol. 24, no. 4, pp. 265–269, 1973. 21

[106] M. M. Kessler, "Bibliographic coupling between scientific papers," *American Documentation*, vol. 14, no. 1, pp. 10–25, 1963. 21

[107] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre, "Fast unfolding of communities in large networks," *J. Stat. Mech.: Theory Exp.*, vol. 2008, no. 10, p. P10008, 2008. 21

[108] M. E. J. Newman, "Modularity and community structure in networks," *Proc. Natl. Acad. Sci.*, vol. 103, no. 23, pp. 8577–8582, 2006. 21

[109] S. Mirjalili, A. H. Gandomi, S. Z. Mirjalili, S. Saremi, H. Faris, and S. M. Mirjalili, "Salp swarm algorithm: A bio-inspired optimizer for engineering design problems," *Adv. Eng. Softw.*, vol. 114, pp. 163–191, 2017. 22

[110] X. Yang, "Bat algorithm for multi-objective optimisation," *Int. J. Bio Inspired Comput.*, vol. 3, no. 5, pp. 267–274, 2011. 22

[111] N. P. Hummon and P. Dereian, "Connectivity in a citation network: The development of DNA theory," *Social Networks*, vol. 11, no. 1, pp. 39–63, 1989. 23

[112] V. Batagelj, "Efcient algorithms for citation network analysis," *CoRR*, vol. abs/0309023, 2003. 23

[113] J. S. Liu and L. Y. Y. Lu, "An integrated approach for main path analysis: Development of the hirsch index as an example," *J. Assoc. Inf. Sci. Technol.*, vol. 63, no. 3, pp. 528–542, 2012. 23

[114] Y. Wang, Y. Tong, and M. Zeng, "Ranking scientific articles by exploiting citations, authors, journals, and time information," in *AAAI: Proc. of the 20th Conference on Artificial Intelligence.* AAAI Press, 2013, pp. 933–939. 23

[115] D. Walker, H. Xie, K.-K. Yan, and S. Maslov, "Ranking scientific publications using a model of network traffic," *J. Stat. Mech.: Theory Exp.*, vol. 2007, p. P06010, 2007. 23

[116] E. Yan, Y. Ding, and C. R. Sugimoto, "P-rank: An indicator measuring prestige in heterogeneous scholarly networks," *J. Assoc. Inf. Sci. Technol.*, vol. 62, no. 3, pp. 467–477, 2011. 23

[117] H. Jain and K. Deb, "An evolutionary many-objective optimization algorithm using reference-point based nondominated sorting approach, part II: handling constraints and extending to an adaptive approach," *IEEE Trans. Evol. Comput.*, vol. 18, no. 4, pp. 602–622, 2014. 23

[118] H. Ishibuchi, Y. Setoguchi, H. Masuda, and Y. Nojima, "Performance of decomposition-based many-objective algorithms strongly depends on pareto front shapes," *IEEE Trans. Evol. Comput.*, vol. 21, no. 2, pp. 169–190, 2017. 23

[119] H. Ishibuchi, Y. Sakane, N. Tsukamoto, and Y. Nojima, "Evolutionary many-objective optimization by NSGA-II and MOEA/D with large populations," in *SMC'09: Proc. of the IEEE International Conference on Systems, Man and Cybernetics.* IEEE, 2009, pp. 1758–1763. 23

[120] S. Milgram, "The small world problem," *Psychol. Today*, vol. 2, no. 1, pp. 60–67, 1967. 25

[121] M. Newman, *Networks: An Introduction.* Oxford University Press, 2010. 25

[122] Y. Lin, C. B. Frey, and L. Wu, "Remote collaboration fuses fewer breakthrough ideas," *Nature*, vol. 623, pp. 987–991, 2023. 25

[123] W. J. Price and L. W. Bass, "Scientific research and the innovative process," *Science*, vol. 164, no. 3881, pp. 802–806, 1969. 25

[124] S. Brin and L. Page, "The anatomy of a large-scale hypertextual web search engine," *Comput. Networks*, vol. 30, no. 1-7, pp. 107–117, 1998. 25

[125] Y. Ding, E. Yan, A. R. Frazho, and J. Caverlee, "Pagerank for ranking authors in co-citation networks," *J. Assoc. Inf. Sci. Technol.*, vol. 60, no. 11, pp. 2229–2243, 2009. 25