A Survey of Evolutionary Multi-Objective Clustering Approaches

CRISTINA Y. MORIMOTO and AURORA POZO, Federal University of Paraná, Brazil
MARCÍLIO C. P. DE SOUTO, University of Orléans, France

This article presents how the studies of the evolutionary multi-objective clustering have been evolving over the years, based on a mapping of the indexed articles in the ACM, IEEE, and Scopus. We present the most relevant approaches considering the high-impact journals and conferences to provide an overview of this study field. We analyzed the algorithms based on the features and components presented in the proposed general architecture of the evolutionary multi-objective clustering. These algorithms were grouped considering common clustering strategies and applications. Furthermore, issues regarding the difficulty in defining appropriate clustering criteria applied to evolutionary multi-objective clustering and the importance of the evolutionary process evaluation to have a clear view of the optimization efficiency are discussed. It is essential to observe these aspects besides specific clustering properties when designing new approaches or selecting/using the existing ones. Finally, we present other potential subjects of future research, in which this article can contribute to newcomers or busy researchers who want to have a wide vision of the field.

CCS Concepts: • Information systems → Clustering; • Computing methodologies → Cluster analysis; • Theory of computation → Evolutionary algorithms; • Applied computing → Multi-criterion optimization and decision-making.

Additional Key Words and Phrases: Multi-objective clustering, Multi-objective optimization, Multi-objective evolutionary algorithms, Multi-criteria clustering

1 INTRODUCTION

The large volume of data generated in recent years has required the improvement of knowledge discovery techniques to facilitate their analysis and understanding. In this context, data clustering approaches have been widely studied and adopted for several purposes, as pattern analysis, decision making, data mining, image segmentation, as well as in other areas such as biology, medicine, and marketing [36, 46, 48, 50].

Traditional clustering algorithms optimize only one clustering criterion and are often very effective for this purpose. However, in general, they may not find all clusters in the datasets with different data structures, neither clusters with shapes hidden in sub-spaces of the original features space. In contrast, the simultaneous optimization of multiple objectives can solve a variety of clustering problems considering different data properties that may improve the robustness of the clustering process and obtain reliable solutions. For example, in Fig. 1, it is shown different data structures, in which the objects with different colors represent different clusters in each subfigure. Algorithms that use the compactness-based criterion, as $k$-means (KM) [66] can detect globular clusters, as shown in Fig. 1(a). The algorithms that take into consideration the density-based criterion, like Shared Nearest Neighbour (SNN) [27], can find clusters with a ring shape, as illustrated in Fig. 1(b). However, SNN cannot find globular clusters, as presented in Fig. 1(a) and KM cannot find the clusters ring shapes, as shown in Fig. 1(b). Moreover, none of them can detect all clusters presented in Fig. 1(c). On the other hand, multi-objective approaches consider two criteria, a compactness-based and a connectedness-based, that can detect all data structures in Fig. 1.

The studies of multi-objective clustering techniques have emerged and increased in the last two decades, exploring the use of multiples criteria to extract patterns and provide multiple partitions as solutions. A variety of Multi-Objective
Clustering (MOC) algorithms has been generated over the years. So, some reviews and surveys have presented an overview of these algorithms: Hruschka et al. [46] introduced a general view of the multi-objective evolutionary clustering in a review of evolutive clustering algorithms; Bong and Rajeswari [8, 9] presented multi-objective clustering trends and methods in image segmentation; Mukhopadhyay et al. [74] presented a survey of multi-objective evolutionary algorithms for data mining, in which some multi-objective clustering algorithms were presented; Mukhopadhyay et al. [79] provided a survey of multi-objective evolutionary clustering, considering a general framework for multi-objective clustering, as a reference guide for those who begin their studies in this area. Gupta and Sharm [36] list some algorithms focused on solving real-life problems. The most recent related work is presented by Khurma and Aljarah [52], which describes some general applications for multi-objective evolutionary clustering methods, providing only one algorithm in their review. The works presented by Bong and Rajeswari [8, 9], and Gupta and Sharm [36] describe algorithms for a specific purpose, while the papers of Hruschka et al. [46] and Mukhopadhyay et al. [74] consider a general context (evolutionary clustering/data mining), describing some multi-objective evolutionary clustering algorithms. In general, these studies, including [52, 79], consider some specific multi-objective approaches, and they do not provide a broad view of the evolutionary multi-objective clustering algorithms.

In this context, this paper provides a survey of Evolutionary Multi-Objective Clustering (EMOC) that considers a mapping of articles in ACM Digital Library, IEEE Xplore and Scopus to provide a comprehensive view of the field. We present the most relevant EMOC algorithms, considering high-impact papers based on the h-index and Scopus percentile scores. These algorithms were grouped by common features or strategies for data clustering. Besides that, we also introduce the most relevant specific applications that researchers have focused on over the years. Furthermore, we introduce a general architecture for evolutionary multi-objective clustering algorithm that provides a general structure of the components and how they are connected, facilitating the understanding of the presented approaches. In particular, we point out some aspects of the multi-objective evolutionary algorithms, like specific categories of algorithms and the analysis/evaluation of the optimizers regarding their application in EMOC approaches. At last, this survey presents the evolution of the EMOC studies and lists some other metaheuristics that have their usage increased over the years.

The remainder of this paper is organized as follows. In Section 2, we present the main concepts concerning clustering. In Section 3, we introduce a general architecture of evolutionary multi-objective clustering. Section 4 presents some numerical data on a set of MOC studies to show the evolution of the publications of the EMOC. Then, in Section 5, we present a general review of the EMOC algorithms, considering algorithms for general purpose and specific application.
Section 6 presents other metaheuristics applied to MOC. Finally, Section 7 highlights our main findings and discusses future works.

2 CLUSTERING

Clustering is a type of unsupervised learning whose goal is to find the underlying structure, composed of clusters present in the data [50]. Objects or observations belonging to each cluster should share some relevant property (similarity) regarding the data domain. Generally, data clustering consists of the decomposition of finite and unlabeled data into subgroups based on similar attributes, or naturally occurring trends, patterns, or relationships in the data [49].

There is not a unique and formal definition of a cluster since the clustering methods and algorithms were proposed for researchers in different fields and applied in a variety of problems and distinct goals. In general, some general properties for the cluster analysis are considered: well-separated clusters, where each object is closer (more similar) to all of the objects in its cluster than to any object in another cluster; connected or contiguous clusters, in which each object is closer to at least one object in its cluster than to any object in another cluster; compact clusters, it represents clusters with small intra-cluster variation, considering the variation between same-cluster data items or between data items and clusters; center-based clusters, in which each object is closer to the center of its cluster than to the center of any other cluster; density-based clusters, where the clusters are regions of high density separated by regions of low density [27, 42, 110].

In general, we can describe the clustering algorithms considering some common aspects. For example, in terms of the clustering process, there are two types of approaches: Crisp Clustering (Exclusive Clustering) — in which each object of a dataset is assigned to a single cluster, and Soft Clustering (Overlapping Clustering) — in which an object may simultaneously belong, to different degrees, to two or more groups of a partition. For example, KM considers a crisp clustering process to minimizes the distance between the objects of a cluster and its center, and Fuzzy C-Means (FCM) [7], a soft KM, minimize the distance from any given data point to a cluster center weighted by the membership of that data point in the cluster. Another aspect is presented by Jain et al. [48, 50], a taxonomy of the traditional clustering approaches based on two general categories: partitional and hierarchical — hierarchical methods produce a nested series of partitions, while partitional methods produce only one. For example, KM, SNN, Minimum Spanning Tree (MST) - clustering [42] and Spectral Clustering (SPC) [101] are partitional algorithms; Single-Linkage (SL) [105], Average-Linkage (AL) [106] and Complete-Linkage (CL) [107] are hierarchical algorithms. Additionally, the number of criteria used and the resulting number of solutions also define other aspects of the clustering algorithms. Traditional clustering algorithms, such as KM, that optimize only one clustering criterion (e.g., compactness) and are often very effective for this purpose. However, they fail for data whose underlying structure obeys a distinct criterion. On the other hand, the Clustering Ensemble (see Section 2.1) can generate a resulting partition considering a set of solutions generated by different clustering algorithms, it helps users overcome the dilemma of selecting an appropriate technique and the corresponding parameters, given a set of data to be investigated [10]. However, the generation of multiple partitions as solutions is a relevant issue in many real scenarios, and the traditional cluster ensemble does not intend to provide multiple partitions [84]. This paper focused on the evolutionary multi-objective clustering approaches that use multiples criteria to extract patterns and provide multiple partitions as solutions, considering both clustering processes (e.g., soft and crisp). This kind of approach can be associated with categorical and/or hierarchical approaches, as well clustering ensemble methods (see Section 2.1), to improve their clustering performance, as presented in Section 5.

Manuscript submitted to ACM
2.1 Clustering Ensemble

The general concept of the clustering ensemble has been applied in the EMOC approaches over the years, that combine different decisions of various clusterings in such a way to achieve the accuracy superior to those of traditional clustering algorithms [10]. Ensemble-based clustering is a modern clustering type of algorithms, defined as follows. Let $X = \{x_1, \ldots, x_n\}$ be a set of $n$ data points, and $\Pi = \{\pi_1, \ldots, \pi_M\}$ be a set of partitions generated by one or more clustering algorithms, a consensus function combines these partitions to obtain the final clustering result $\pi^*$, and to improve the quality of the clustering results [28]. There are a variety of clustering ensemble algorithms, considering different strategies and applications, as shown in [10]. Details of the clustering ensemble usage in EMOC are presented in Section 3.4 and 3.5.

2.2 Clustering Validation

The clustering approaches are evaluated regarding Clustering Validity Indices (CVIs), which define how well a partition fits the structure underlying the data. There are three types criteria [12]: relative, internal and external. Relative criteria are based on comparisons of partitions generated by the same algorithm with different parameters or different subsets of the data. Internal criteria refer to quality measures based on calculating properties of the resulting clusters, establishing the validity of a cluster-based exclusively in the dataset itself, for example, how much a cluster is justified by means of the proximity matrix. External criteria lie in prior knowledge of structures in the dataset to evaluate the given partitions generated by an algorithm in contrast with a model partition or labeled data, denominated True Partition[^1], provided by specialists. We detail the CVIs and their application in EMOC approaches in Section 3.3.

3 A GENERAL ARCHITECTURE OF EMOC ALGORITHM

In multi-objective optimization the goal is to find a vector of decision variables, $\pi$, which satisfies the $p$ inequality constraints, Equation (1), the $q$ equality constraints, Equation (2), and optimizes the vector $F(\pi)$ of $z$ objective functions, Equation (3) [56]. In this context, the constraints define the feasible region of acceptable solutions, that must be fulfilled while optimizing objective functions. In other words, the Multi-objective Optimization Problem (MOP) can be described as the minimization (or maximization) of the vector function $F(\pi)$ that maps a tuple of parameters decision variables to a tuple of objective, where $z \geq 2$ [134].

\[
g_i(\pi) \leq 0, \quad i = \{1, \ldots, p\} \quad (1)
\]

\[
h_j(\pi) = 0, \quad j = \{1, \ldots, q\} \quad (2)
\]

\[
F(\pi) = (f_1(\pi), f_2(\pi), \ldots, f_z(\pi)) \quad (3)
\]

Multiples solutions arise in even the simplest non-trivial case of two objectives competing objectives. As the number of competing objectives increases and less well-behaved objectives are considered, the problem rapidly becomes increasingly complex. In this context, the Evolutionary Algorithms (EAs) are considered well-suitable to multi-objective optimization because they address both search and multi-objective decision making (while some approaches focus on search and others on multi-criteria decision making), also can search partially ordered spaces for several alternative trade-offs. EA uses a heuristic solution-search or optimization technique based on the principle of evolution through selection [134].

[^1]: The True Partition or ground truth is the labeled data that form the real partition, the underlying structure of the data.
Fig. 2. A general architecture of Evolutionary Multi-objective Clustering

In the context of the clustering problem, Fig. 2 presents a general architecture of an evolutionary multi-objective clustering that is composed of three general phases: initialization, optimization, and selection. This architecture considers that given a dataset $X = \{x_1, x_2, \ldots, x_n\}$, in the initialization phase, traditional data clustering algorithms (or random generator methods) are performed to build the individuals that compose the initial population. Each individual of this population is a clustering solution with a specific encoding or representation. The representation choice has an important role in multi-objective evolutionary clustering because it can affect the generation of new solutions in the evolutionary process. In the optimization phase, the initial population is taken as input to the evolutionary optimization to iteratively generate a final population. The main aspects of the optimization phase are the objective functions, the selection procedure and particularities of the evolutionary multi-objective approach, the crossover, and mutation operators. In general, EMOC approaches consider the clustering criteria as objective functions. In evolutionary optimization flow, the crossover and mutation operators allow exploring other feasible solutions, and the selection procedure supported by a sorting mechanism contributes to convergence of the Final Population. Finally, the selection phase is applied to determine the final set of solutions, in which the solutions in the Final Population are filtered according to prior criteria, considering a suitable number of solutions, $s'$, to be presented to the data experts. In the following, we present more details regarding the main concepts and elements of evolutionary multi-objective clustering.

3.1 Solution Representation and Initialization

The choice of the solution representation (or encoding) must consider the information necessary to be manipulated by the evolutionary operators to generate new feasible solutions. In general, the most popular types of clustering representation solutions are: (i) Label-based representation, that considers labels for each object in the partition; usually, the length of an encoding of the solution is equal to the number of objects in the dataset, and each position denotes the cluster label of the respective object; (ii) Prototype-based encoding, usually applied in centered-based clustering, in which cluster prototypes (such as centroids, medoids, or modes) are encoded in solution representation. (iii) Locus-based adjacency graph representation corresponds to a graph containing a vertex for each data point, and the links between two data points represent the edges. The linked objects represent the clusters in the solution.
In particular, some approaches use a binary representation to define the labels or prototypes instead of using numerical values. In [98, 99], each chromosome includes \( n \times k \) bits, and each reserved \( k \) bits provide the cluster number of the corresponding instance. In [91], each data point is a candidate center, and a binary encoding is applied to define whether a data point is a center or not. Besides that, it is possible to consider other aspects of the clustering problem in the representation. For example, in [18], FCM parameters and features weights are applied to represent the solution. In [120, 122, 129], they use the center information associated with a (center) weight to encode the solutions. In [19, 130], they consider an input as a linear combination of base elements (e.g., parameters or coefficients), which are chosen from an over-complete dictionary to design the sparse-based representation.

Regarding the initialization, a common practice is to use a random method or traditional clustering algorithms to generate the initial population. In particular, most prototype-based encoding approaches use a random method in the initialization. On the other hand, the label-based encoding takes advantage of not requiring decode of the solutions, turning possible to apply most of the traditional clustering algorithms in the initialization. At last, the adjacency graph representation can rely on a graph-based method in the initialization, as MST-clustering, taking advantage of its data structure.

### 3.2 Multi-objective Evolutionary Optimization

In general, the EMOC algorithms rely on general-purpose Multi-objective Evolutionary Algorithms in the Optimization phase. The choice of the multi-objective approach must consider the number of objectives functions and the characteristics of the application, in which it is possible to explore some aspects, as user preference, diversity of the solutions, among other features. The most traditional category of the multi-objective algorithms is Pareto-based, where the solutions are evaluated and compared, considering the Pareto dominance.

More specifically, given two candidates solutions \( \pi_i \) and \( \pi_j \), \( \pi_i \) dominates \( \pi_j \) (denoted as \( \pi_i \prec \pi_j \)), if and only if: 1) \( \pi_i \) is strictly better than \( \pi_j \) in at least one of all objectives considered, and 2) \( \pi_i \) is not worse than \( \pi_j \) in any of the objectives considered. In this context, the goal of the optimization process is to find the set of all non-dominated solutions, that is, the Pareto-optimal front (PF) \( \Pi_F = \{ \pi_1, \ldots, \pi_{|\Pi_F|} \} \). For example, Fig. 3 shows a Pareto set of two objective functions that should be minimized. Points A and B are the non-dominated solutions and hence lie on the Pareto front. Point C is dominated by points A and B, so it does not lie on the frontier. Due to their population-based nature, evolutionary algorithms are able to approximate the whole PF of a given multi-objective problem in a single run. Consequently, they have been a popular choice for the design of multi-objective data clustering techniques [46, 79]. In this context, the Multi-objective Evolutionary Algorithms (MOEA) are applied to solve a MOP with \( z \geq 2 \). However, the traditional techniques based on the Pareto dominance have their effectiveness degrade (convergence and diversity difficulties) when applied in problems with more than three objectives, and the computational complexity of non-dominated sorting considerably increases. The Many-Objective Evolutionary Algorithms (MaOEA) have been proposed to deal with this scalability issue, in which the Many Objectives Problem can be defined as a MOP with \( z \geq 4 \) [56].

Regarding the Pareto-based approaches, NPGA - Niched Pareto Genetic Algorithm [45] is designed along with the natural analogy of evolution of distinct species exploiting different niches or resources in the environment, in which the main strategy rely on the tournament selection among a population’s individuals and Pareto dominance. The PESA-II - Pareto Envelop-based Selection Algorithm version 2 [14] is an elitist method, where the diversity mechanism is cell-based density. The NSGA-II - Non-dominated Sorting Genetic Algorithm version II [15] is an elitism method that employs a ranking based on non-domination sorting associated with crowding distance. The SPEA-2 - Strength Pareto
Evolutionary Algorithm version 2 [132] also is an elitism method that applies the concept of the strength of dominators as a fitness assignment, employing a density based on the $k$th nearest neighbor to preserve the diversity. In particular, Dutta et al. [22–24] consider a specific MOEA that uses the Pareto-based selection and an intermediate population to build the next generation. Besides the solutions on the front, for every odd generation or if for any generation the previous generation does not produce Pareto solutions greater than 2, then the population is created by the initialization procedure. This introduces new chromosomes in the population and induces diversity in the population.

Beyond that, Li et al [56] defined other categories considering other criteria to evaluated and compared the solutions in MOEAs/MaOEAs: (a) Relaxed dominance-based, that use a variant of dominance, as value-based (that change the objective values by modifying the Pareto dominance of the solutions when comparing them) or number-based dominance (that compare a solution to another by counting the number of objectives where it is better than, the same as, or worse than the other); (b) Diversity-based, that apply a customized diversity-based approach, for example, the SDE (Shift-Based Density Estimation) where the diversity is taken as the first criterion instead of the convergency, it is possible because SDE shifts the positions of the solutions to measure the density of the neighborhood of the solution, allowing both the distribution and the convergence information to be used in the comparison of the solutions; (c) Aggregation-based, that apply aggregation functions to evaluate the solutions, that can be divided into two: aggregation of objective values and aggregation of objective ranks; (d) Indicator-based, that aim to maximize the value of a specific indicator, that can be divided into three classes: hypervolume driven, distance-based indicator driven, and R2 indicator driven; (e) Preference set-based, in which the user’s preference is considered in the optimization process, that can be divided into three classes based on the timing of the set of preferences is used: a priori (selection before the search), a posteriori (selection after the search), a interactive (selection during the search), a a posteriori (selection after the search); (f) Reference-based, that considers a set of reference solutions which is applied to measure the quality of the solutions and guide the search during the evolutionary optimization process, as NSGA-III [16] and RVEA [13]; (g) Dimensionality reduction that seeks to simplify the problem reducing their complexity, where the number of objectives can be reduced gradually during the search process (online) or the dimensionality reduction is carried out after obtaining a set of Pareto-optimal solutions (offline) [56]. Additionally, it is possible to consider another category, a Hybrid-based, that combine two or more approaches to overcome their particular problems, for example, the MOEA/DD - Multi-Objective Evolutionary Algorithm based on Dominance and Decomposition approaches [59] combines two categories of strategies Pareto dominance and aggregation.

Besides the MOEA general strategies above-mentioned, in general, the optimization in EMOC differs in objectives functions and evolutionary operators, presented in the following.
3.3 Objective Functions

The objectives functions represent the criteria or features that must be optimized. In general, CVIs that consider internal and relative criteria are used as clustering objectives functions. Objective functions in MOPs must be conflicting with one another (i.e., solutions in which one objective cannot be improved without worsening another one) to obtain a set of non-dominated solutions rather a single optimal solution [116]. There is a variety of CVIs - Clustering Validity Indices used as objective functions. It includes traditional clustering criteria and objective functions designed for the EMOCl. In the following, we present the objective functions applied in the most relevant EMOC approaches, introduced in Section 5.

3.3.1 Compactness criteria - based on intra-cluster measures. In this sub-section, we present the objective functions that consider the compactness of the clusters to be optimized: average within group sum of squares, overall deviation, K-Mode internal distance, K-Mode weighted internal distance, intra-cluster entropy, homogeneity, intra-cluster variance and total within-cluster variance for crisp clustering, and fuzzy compactness for soft clustering.

The Average Within Group Sum of Squares (AWGSS) is computed by the average of the distance between each object in the cluster and its centroid, as present in Equation (4), where \( k \) is the number of clusters in the partition \( \pi \), \( n_i \) denotes the number of the points in the \( i \)th cluster, and \( d(.,.) \) is the selected distance function applied to compute the distance between each object \( x \) in the cluster \( c_i \) and its centroid \( z_i \). It must be minimized to obtain compact clusters [53].

\[
AWGSS(\pi) = \frac{1}{n} \sum_{i=1}^{k} \sum_{x \in c_i} d(x, z_i) \tag{4}
\]

The overall Deviation (Dev) is computed as the overall summed distances between data points and their corresponding cluster center, as defined in the Equation (5), where \( \pi \) denotes a partition, \( x_i \) is an object that belongs to the cluster \( c_k \), \( z_k \) is the centroid of cluster \( c_k \), and \( d(.,.) \) is the selected distance function. Dev must be minimized in order to obtain compact clusters [39].

\[
Dev(\pi) = \sum_{c_k \in \pi} \sum_{x_i \in c_k} d(x_i, z_k) \tag{5}
\]

Sert et al. [98, 99] considered the K-Mode internal distance (Km_{id}) and K-Mode weighted internal distance (Km_{wtd}) as objective functions. The K-mode internal measures are computed in a similar way to Dev, in which it is applied mode instead of the centroid. Km_{id} and Km_{wtd} should be minimized as objective.

The intra-cluster Entropy (Entropy) is used to measure the degree of similarity between each cluster center and data objects belongs to that cluster, as the probability of grouping all the data objects into that particular cluster. A larger value of this index implies a better clustering [91–93].

\[
Entropy(\pi) = \frac{1}{n} \sum_{i=1}^{k} \left[ 1 - \left( g(z_i) \log_2 g(z_i) + (1 - g(z_i)) \log_2 (1 - g(z_i)) \right) \right] \frac{1}{k},
\]

where \( g(z_i) = \frac{1}{n} \sum_{j=1}^{n} \left( 0.5 + \frac{\cos(z_i, x_j)}{2} \right) \sum_{j=1}^{n} \cos(z_i, x_j) = \frac{\left( \sum_{r=1}^{f} z_{ir} x_{jr} \right)}{\sqrt{\sum_{r=1}^{f} z_{ir}^2} \sqrt{\sum_{r=1}^{f} x_{jr}^2}} \) \tag{6}

\( k \) is the number of clusters in the partition \( \pi \), \( n \) is the number of objects in the dataset, \( f \) is the size of the dimensional feature space, \( x_j \) is an object that belongs to cluster \( c_i \), \( z_i \) is the centroid of the cluster \( c_i \), \( g(z_i) \) is average similarity between the cluster center \( z_i \) and the data object belong to cluster \( c_i \), \( \cos(.,.) \) denote the cosine distance.
The **Homogeneity** ($H$) is computed by the sum of the average minimal intra-cluster distance, according to Equation (7), where $\pi$ denotes a partition, $n_i$ is the number of objects in the cluster $c_i$, $\min(d(x_j, x_j))$ is the lowest distance between the points $x_j$ in the cluster $c_i$ and the cluster mode $m_i$, and $d(\cdot, \cdot)$ is the selected distance function. The $H$ must be maximized to obtain homogeneous clusters [21].

$$H(\pi) = \sum_{i=1}^{k} \left( \frac{\sum_{j=1}^{n_i} \min(d(m_i, x_j))}{n_i} \right)$$  \tag{7}

The **intra-cluster Variance** ($Var$) is conceptually similar to Dev, as shown in Equation 8, where $k$ is the number of cluster in the partition $\pi$, $n$ is the number of objects in the dataset, $x_i$ is an object that belongs to the cluster $c_k$, $z_k$ is the centroid of the cluster $c_k$, and $d(\cdot, \cdot)$ is the selected distance function. It also must be minimized to obtain compact clusters [35].

$$Var(\pi) = \frac{1}{n} \sum_{c_k \in \pi} \sum_{x_i \in c_k} d(x_i, z_k)$$  \tag{8}

The **Total Within-Cluster Variance** ($TWCV$) also is applied to identify sets of compact clusters, as defined in Equation (9). The goal is to minimize $TWCV$ to obtain compact clusters [20].

$$TWCV(\pi) = \sum_{i=1}^{n} \sum_{j=1}^{f} w_{ij} \sum_{r=1}^{m} (x_{ir} - z_{jr})^2,$$

where $z_{jr} = \frac{\sum_{i=1}^{n} w_{ij} x_{ir}}{\sum_{i=1}^{n} w_{ij}}$, and $w_{ij}$ is defined as:

$$w_{ij} = \begin{cases} 1, & \text{if } i^{th} \text{ object belongs to } j^{th} \text{ cluster} \\ 0, & \text{otherwise,} \end{cases}$$

$k$ denotes the total number of clusters in the partition $\pi$, $n$ is the number of objects in the dataset, $f$ is the number the features, $x_{ir}$ denotes the $r$th feature value of the $i$th data point, $z_{jr}$ is the centroid of the $j$th cluster of the $r$th features, and $w_{ij}$ represent the data points that belongs to the $j$th cluster, where $w_{ij} \in \{0, 1\}$ and $\sum_{i=1}^{n} w_{ij} = 1$.

The **Fuzzy Compactness** ($J_m$) represents the global fuzzy cluster variance, as defined in Equation (10), where $\pi$ denotes a partition, $x_j$ is an object that belongs to the cluster $c_l$, $z_l$ is the centroid of cluster $c_l$, and $d(\cdot, \cdot)$ is the selected distance function, $u_{ij}$ is membership degree of the $j$th data point to the $i$th cluster, and $m$ is the fuzzy exponent. The smaller value of $J_m$ corresponds to more compact clusters [7].

$$J_m = \sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m d(z_l, x_j)$$  \tag{10}

Zhu et al. [129] introduced an adapted $J_m$ that considers the cluster weighting subspace, the **Fuzzy weighting subspace clustering** ($J_{wm}$), as defined in the Equation (11), where $k$ is the number of clusters, $n$ is the number of the objects in the dataset, $f$ is the number attributes (or vector of features), and the $w_{ir}$ is defined in Equation (12).

$$J_{wm} = \sum_{i=1}^{k} \sum_{j=1}^{n} \sum_{r=1}^{f} w_{ir} u_{ir}^m d(x_{jr} - z_{ir})^2$$  \tag{11}

$$w_{ir} = \left( \frac{\sum_{j=1}^{n} u_{ij}^m d(x_{jr} - z_{ir})^2}{\sum_{r=1}^{f} \sum_{j=1}^{n} u_{ij}^m d(x_{jr} - z_{ir})^2} \right)^{1/r-1},$$

where

$$u_{ij} = \frac{(\sum_{r=1}^{f} w_{ir}^m d(x_{jr} - z_{ir})^2)^{-1/m-1}}{\sum_{i=1}^{k} (\sum_{r=1}^{f} w_{ir}^m d(x_{jr} - z_{ir})^2)^{-1/m-1}}$$  \tag{12}
\(x_j, r\) denote rth object that belongs to the cluster \(c_j, z_i\) is the centroid of cluster \(c_i, d(., .)\) is the selected distance function, \(m\) is the fuzziness exponent and \(\tau\) is the fuzzy weighting index. \(J_{wm}\) must be minimized to improve the clustering.

### 3.3.2 Connectedness criteria - based on neighborhood relationship.

Here, we present objective functions that consider the connectedness of the clusters to be optimized: connectivity index, connectivity based on Person Correlation, and data continuity degree. All of these criteria are related to crisp clustering.

The **Connectivity** \((\text{Con})\) index [39] is computed according to Equation (13), where \(\pi\) is a partition, \(n\) is the number of objects, \(L\) is the parameter that determines the number of nearest neighbors that contributes to the connectivity, \(nnij\) is the \(j\)th nearest neighbor of object \(x_i, c_k\) is a cluster that belongs to \(\pi\), and \(d(., .)\) is the selected distance function. \(\text{Con}\) as objectives must be minimized.

\[
\text{Con} (\pi) = \sum_{i=1}^{n} \sum_{j=1}^{L} f(x_i, nnij), \text{where } f(x_i, nnij) = \begin{cases} \frac{1}{j}, & \text{if } \hat{p}c_k : x_i, nnij \in c_k \\ 0, & \text{otherwise} \end{cases} \tag{13}
\]

The **Data Continuity Degree** (DCD) measures the continuity of the data as a graph structure inside the clusters, in which in connectivity factor is computed as a sum of the number of the MST edges, considering all nodes connected within the neighborhood related to each node, due to the graph is not fully-connected, this process is repeated with each connected component. The arithmetic average value of the metric is the result of this objective [69].

### 3.3.3 Separation criteria - based on inter-cluster measures.

In this sub-section, we present the 8 objective functions that consider the separation of the clusters, six of them, the average between-group sum of squares, inter-cluster distance, K-Mode external distance, K-Mode weighted external distance, separation index, and graph-based separation, are CVIs for crisp clustering, and the other two, fuzzy separation and fuzzy overlap separation, are CVIs for overlap clustering.

The **Average Between-Group Sum of Squares** (ABGSS) is computed as the average distance between the clusters’ centroids and the centroid of the data, as defined in Equation (14), where \(k\) is the number of clusters in the partition \(\pi, n_i\) denotes the number of the points in cluster \(c_i, d(., .)\) is the selected distance function applied to compute the distance between each cluster centroid \(z_i\) and the dataset center \(\bar{z}\). It must be maximized to obtain well-separated clusters [53].

\[
\text{ABGSS}(\pi) = \frac{\sum_{i=1}^{k} n_i d(z_i, \bar{z})}{k} \tag{14}
\]

The inter-cluster distance **Average Separation** \((\text{Sep}_{AL})\) is computed according to Equation (15), it measures the average separation distance between all clusters. \(\text{Sep}_{AL}\) must be maximized obtain better clustering [91].

\[
\text{Sep}_{AL}(\pi) = \frac{1}{k(k - 1)/2} \sum_{i<j}^{k} d(z_i, z_j), \tag{15}
\]

where \(k\) is the number of clusters in the partition \(\pi, z_i\) and \(z_j\) are the cluster centers of two distinct clusters, and \(d(., .)\) is the selected distance function.

Sert et al. [98, 99] introduce the use the **K-Mode external distance** (Km\(_{ed}\)) and **K-Mode weighted external distance** (Km\(_{wed}\)) as objective functions. These measures are similar to \(\text{Sep}_{AL}\), however considering the mode instead the centroid. Km\(_{ed}\) and Km\(_{wed}\) should be maximized as objective function.

The **Separation Index** \((\text{Sep}_{CL})\) is computed by the sum of the distance between every two tuples (data points) in different clusters, as shown in Equation (16), where \(x_i\) and \(x_j\) are objects in different clusters and \(d(., .)\) is the selected distance function.
distance function. It must be maximized to get well-separated clusters [22].

\[
Sep_{CL}(\pi) = \sum_{i,j=1, i \neq j}^{n} d(x_i, x_j)
\]  

The Graph-based Separation (Sep\textsubscript{graph}) is computed by the arithmetic average value of the edge weights between the different clusters, as shown in Equation (17), where \( c \) is a cluster, \( G \) is the \( K \text{size-Graph} \), \( v_i \) is the vertex \( i \), \( w_{ij} \) is the edge weight value from node \( i \) to node \( j \). Sep\textsubscript{graph} must be maximized to improve the cluster separation [69].

\[
Sep_{graph} = \left( \sum_{v_j \in G} \frac{w_{ij} | v_j \notin c}{G - c} \right) / c
\]  

The Fuzzy Separation [75] is computed according to Equation (18), where the fuzzy membership is defined by \( \mu_{ij} \), \( d(z_i, z_j) \) is the distance between two centroids \( z_i \) and \( z_j \), in which \( d(., .) \) is the selected distance function. To get well-separated clusters, the \( Sep_{fuzzy} \) must be maximized.

\[
Sep_{fuzzy} = \sum_{i,j=1, i \neq j}^{n} \mu_{ij}^m d(z_i, z_j), \text{ where } \mu_{ij} = 2 / \left( \sum_{i=1}^{k} \left( \frac{d(z_j, z_i)}{d(z_j, z_i)} \right)^{1/(m-1)} \right), i \neq j
\]  

The Fuzzy Overlap Separation (Sep\textsubscript{fuzzy}) considers the combination of the \( l \)-order overlap and inter-cluster separation, composed of a \( t \)-normal function \( \top \) and \( t \)-conorm \( \bot \) to formulate the Fuzzy Overlap Separation [82, 119]. Sep\textsubscript{fuzzy} is defined in Equation (19) as follows:

\[
Sep_{fuzzy} = \frac{1}{n} \sum_{i=1}^{n} \frac{O \top (u_i(x_i), k)}{\max_{j=1, k} u_{ij}}
\]

where \( n \) is the number of objects in the dataset, \( k \) is the number of clusters in the partition \( \pi \), \( u_{ij} \) is membership degree of the \( j \)th data point to the \( i \)th cluster, \( O \top (u_i(x_i), k) \) is the overlapping degree that considers triplets of clusters up to a \( k \)-tuple of clusters combinations. Sep\textsubscript{fuzzy} measures the isolation of clusters, which is preferred to be large.

### 3.3.4 Separation and Compactness criteria - based on inter and intra-cluster measures

In this sub-section, we present CVIs designed to obtain well-separated and compact clusters that consider the relationship between the intra-cluster and inter-clusters aspects. Here, we present six CVIs for crisp clustering (categorical data clustering with subjective factors, Calinski-Harabasz, Davies-Bouldin, Dunn, modularity, and silhouette) and four CVIs for soft clustering (\( I \), addition feature weight, Xeni-Beni, soft subspace Xie-Beni).

The Categorical Data Clustering with Subjective factors (CDCS) index is computed by the ratio of the intra-cluster cohesion and inter-cluster similarity for the categorical data clustering, as shown in Equation (20) [130].

\[
CDCS = \frac{\text{intra}}{\text{inter}} = \sum_{r=1}^{k} \frac{\sum_{s=1}^{n} \frac{1}{f} \left( \max_{r=1}^{n} P(A_r = a_r^i | c_i) \right)^3}{\sum_{p=1}^{k} \sum_{q=1}^{k} S(c_p, c_q)^{1/2} [c_p \cap c_q]} \frac{1}{(k-1)n},
\]

where \( S(c_p, c_q) = \sum_{r=1}^{f} \sum_{i=1}^{l_r} \min \left( P(A_r = a_r^i | c_p), P(A_r = a_r^i | c_q) + \epsilon \right) \),
\(A_r\) is a set of attribute values, \(a_r\) denotes the number of attribute values for the \(r\)th attribute, \(P(A_r = a_r | c_j)\) is the probability of \(a_r\) for the \(r\)th attribute in cluster \(c_j\), \(S(c_p, c_q)\) is a similarity of two clusters, \(k\) is the number of clusters in the partition \(\pi\), \(n\) is the number of objects in the dataset, and \(\varepsilon\) is a small value in case that each component is 0.

The **Calinski-Harabasz (CH)** index is based on the degree of dispersion between clusters. It can take values in \([0, \infty]\) with higher values indicating better clustering. CH is computed by the ratio of the sum of between-clusters dispersion and inter-cluster dispersion for all clusters, as defined in Equation (21), where \(k\) is the number of clusters in the partition \(\pi\), \(z_i\) is the centroid of the \(i\)th cluster, \(z\) is the center of the dataset, \(n_i\) denotes the number of the points in cluster \(c_i\), and \(d(., .)\) is the selected distance function [131].

\[
CH(\pi) = \frac{\sum_{i=1}^{k} n_i d(z_i, z)}{(n - k) \sum_{i=1}^{k} \sum_{x \in c_i} d(x, z_i)} (k - 1)
\] (21)

The **Davies-Bouldin (DB)** index is computed as the ratio of the sum of within-cluster scatter to between-cluster separation (\(R_i\)), as defined in Equation (22) [19, 24, 112, 131].

\[
DB(\pi) = \frac{1}{k} \sum_{i=1}^{k} R_i, \text{ where } R_i = \max_{j \neq i} \left( \frac{S_i + S_j}{d(z_i, z_j)} \right), \text{ and } S_i = \frac{1}{|n_i|} \sum_{x \in c_i} d(x, z_i),
\] (22)

\(k\) is the number of clusters in the partition \(\pi\), \(S_i\) is the scatter within the \(i\)th cluster, \(z_i\) and \(z_j\) are centroids of two distinct clusters \(c_i\) and \(c_j\), \(n_i\) denotes the number of the points in cluster \(c_i\), \(x\) is an object that belongs to the cluster \(c_i\), and \(d(., .)\) is the selected distance function. The minimum value of this DB is zero, with lower values indicating a better clustering.

The **Dunn** index is calculated by the ratio between the minimum inter-cluster distance \(\delta(c_i, c_j)\) to the maximum cluster diameter (the maximum intra-cluster distance) \(\max \Delta(c_i)\), as shown in Equation (23) [24, 131].

\[
Dunn(\pi) = \min_{1 \leq i < j \leq k} \left( \min_{1 \leq i < j \leq k} \left( \frac{\delta(c_i, c_j)}{\max \Delta(c_i)} \right) \right),
\] (23)

where \(\delta(c_i, c_j) = \min_{x_a \in c_i, x_b \in c_j} \{d(x_a, x_b)\}\), and \(\Delta(c_i) = \max_{x_a, x_b \in c_i} \{d(x_a, x_b)\}\), \(c_i\) and \(c_j\) are distinct clusters in the partition \(\pi\), and \(d(., .)\) is the selected distance function applied to compute the distance between two objects \(x_a\) and \(x_b\). The Dunn index has a value between zero and infinity and it should be maximized to obtain well-separated and compact clusters.

The **Modularity** is based on the modularity method applied to detect network structures. It is computed as the relationship of the sum of distances of the objects in the same cluster \(c_i\) and the sum of distances considering the objects in the dataset \(X\), as defined in Equation (24), where \(c_i\) is the \(i\)th cluster in the partition \(\pi\), and \(d(., .)\) is the selected distance function applied to compute the distance between two objects \(x_i\) and \(x_j\) [61].

\[
Modularity(\pi) = \sum_{i=1}^{k} \left( \frac{\sum_{x_j, x_i \in c_i} d(x_j, x_i)}{\sum_{x_j, x_i \in X} d(x_j, x_i)} - \left( \frac{\sum_{x_j, x_i \in X} d(x_j, x_i)}{\sum_{x_j, x_i \in X} d(x_j, x_i)} \right)^2 \right)
\] (24)

The **Silhouette (Sil)** is computed using the mean intra-cluster distance and the mean nearest-cluster distance for each sample, as shown in Equation (25). Sil produces values between \(-1\) and \(1\), a higher value corresponds to a better
is the number of clusters in the partition, weighted coverage density indices consider the relation of the occurrence of the objects in a categorical dataset. The within-cluster compactness \( J_i \) is the number of clusters in the partition, \( \pi \), \( z_i \) and \( z_j \) are centroids of distinct clusters \( c_i \) and \( c_j \), \( x_j \) is the \( j \)th object in the cluster \( c_j \), \( d(.,.) \) is the selected distance function, and \( u_{ij} \) is membership degree of the \( j \)th data point to the \( i \)th cluster. A larger value of index \( I \) implies better clustering.

The Addition feature weight \( (J_{add}) \) is applied to minimize the information of the negative weight entropy and the separation between clusters, as defined in Equation (27). It is composed by \( \text{Sep}_i \), that is computed according to Equation (18), \( \sigma \) a preset value that prevents the denominator from becoming zero, and \( A_w \) denotes the average value of the important weights, which are more than or equal to the mean value \( (1/f) \) for the \( i \)th cluster [129].

\[
J_{add} = \sum_{i=1}^{k} \left( \frac{A_w}{(\text{Sep}_i + \sigma)} + \sum_{f=1}^{f} \frac{w_{ir} \log w_{ir}}{\delta_k} \right), \quad \text{where} \quad A_w = \sum_{f=1}^{f} \frac{\delta_r w_{ir}}{\delta_k}, \quad \delta_r = \begin{cases} 1, & \text{if } w_{ir} > 1/f \\ 0, & \text{otherwise} \end{cases}
\]

where \( k \) is the number of clusters in the partition, \( f \) is the number of attributes, and \( w_{ir} \) takes the value in \([0, 1]\), which corresponds to a soft partition of features.

The Xeni-Beny (XB) index is defined as a function of the ratio of the total fuzzy cluster variance \( (J_m) \) to the minimum separation of the clusters \( (\text{Sep}) \), as defined in Equation (28), where \( k \) is the number of clusters in the partition \( \pi \), \( z_i \) and \( z_j \) are centroids of distinct clusters \( c_i \) and \( c_j \), \( x_j \) is the \( j \)th object in the cluster \( c_j \), \( d(.,.) \) is the selected distance function, \( u_{ij} \) is membership degree of the \( j \)th data point to the \( i \)th cluster, and \( m \) is the fuzzy exponent. It must be minimized to obtain well-separated and compact clusters [18, 130].

\[
\text{XB}(\pi) = \frac{J_m}{n \times \text{sep}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^m d(z_i, x_j)}{n \times \min_{i \neq j} \{d(z_i, z_j)\}},
\]

The Soft Subspace Xie-Beni (SSXB) index was extended from the XB. It is defined as the ratio of the fuzzy weighting within-cluster compactness \( (J_{swm}) \) to the fuzzy minimum weighting between-cluster separation \( (J_{wsep}) \), as defined in Equation (29), where \( n \) is the number of objects in the dataset, \( k \) is the number of clusters in the partition, \( f \) is the number of attributes, \( d(.,.) \) is the selected distance function, and the \( w_{ir} \) and \( u_{ij} \) are defined in Equation (12). SSXB must be minimized as an objective function [129].

\[
\text{SSBX} = \frac{J_{swm}}{n \times J_{wsep}} = \frac{\sum_{i=1}^{k} \sum_{j=1}^{n} u_{ij}^2 \sum_{r=1}^{f} w_{ir}^2 d(x_{jr} - z_{jr})^2}{n \times (\min_{i \neq j} \{ \sum_{r=1}^{f} w_{ir}^2 d(z_{ir} - z_{jr})^2 + \sum_{r=1}^{f} w_{ir}^2 d(z_{ir} - z_{jr})^2 / 2 \})}
\]

3.3.5 Other criteria. Here, we present the other criteria applied as objective functions. Cluster cardinality and expected weighted coverage density indices consider the relation of the occurrence of the objects in a categorical dataset. The similarity index is the only relative CVI that compares partitions used as the objective function, while the other CVIs
consider the data properties of each partition. The sparsity and reconstruction error are two particular objective
functions designed for spectral clustering.

The Cluster Cardinality Index (CCI) considers a set of operations to describe the property and structure of
categorical data [130]. It is computed according to Equation (30)

\[
CCI = \frac{1}{k} \sum_{i=1}^{k} \max_{l, i = 1, l \neq i} \left( \frac{CI(i) + CI(l)}{CI(i, l)} \right),
\]

(30)

where \( CI(i) = \frac{1}{f} \sum_{t=1}^{f} \frac{|A_{tr}|}{c_i}, CI(i, l) = \frac{1}{f} \sum_{t=1}^{f} \frac{|A_{tr} \cap A_{lr}| - |A_{tr} \cup A_{lr}| + 1}{|A_{tr} \cap A_{lr}|} + 1, \)

\(A_{tr}\) and \(A_{lr}\) are the set of categorical values of \(r\)th attribute within cluster \(c_i\) and \(c_j\). A larger value of CCI implies better
clustering.

The intra-cluster Expected Weighted Coverage Density (EWCD) considers the relation of the objects in a trans-
rental dataset. The transrental dataset is composed by \(n\) transactions considering the set of items \(I = \{I_1, I_2, \ldots, I_m\}\),
where the transaction \(t_j (1 \leq j \leq n)\) is a set of items \(t_j = \{I_{j1}, I_{j2}, \ldots, I_{jm}\}\), such that \(t_j \subseteq I\). In this context, the
WCD-Weighted Coverage Density of one cluster is defined as the sum occurrences of all items in a cluster divided by
the number of distinct items and the total number of items in this cluster. So, the EWCD of the partition \(\pi\) is defined as
an average sum of the WCD in all clusters, as presented in the Equation (31), where \(n\) is the number of objects in the
dataset, \(k\) is the number of clusters in the partition \(\pi\), \(n_j\) is the number of elements in \(c_i\), \(I_{ij}\) is the \(j\)th item set in the
cluster \(c_i\), \(\text{occur}(I_{ij})\) define the number of occurence of the \(j\)th item in cluster \(c_i\), and \(S_i\) is the sum occurrences of all
items in cluster \(c_i\) [98, 99].

\[
EWCD(\pi) = \frac{1}{n} \sum_{i=1}^{n} \frac{n_i \cdot WCD}{n} = \frac{1}{n} \sum_{i=1}^{k} \left( \frac{\sum_{j=1}^{n_i} \text{occur}(I_{ij})^2}{S_i} \right).
\]

(31)

Li et al. [58] provide the Similarity (Sim) index, that evaluates the similarity of one partition to others with a
similarity matrix, as defined in Equation (32). Where \(\text{similarity}(\pi_i, \pi_j)\) is the value of Adjusted Rand Index (ARI),
between the partition \(\pi_i\) and \(\pi_j\). The similarity can be used to evaluate the diversity of the solutions in an evolutive
strategy. This index should be minimized as an objective [58].

\[
Sim = \frac{1}{n} \sum_{j=1}^{n} \text{similarity}(\pi_i, \pi_j),
\]

(32)

The Adjusted Rand Index (ARI) [90], a corrected-for-chance version of the Rand index [47], computes the probability
of two objects of two partitions belong to the same cluster or different clusters, as defined in Equation (33), where \(n_{ij}\) is
the number of common objects between the clusters \(c_i\) in \(\pi_a\) and \(c_j\) in \(\pi_b\), \(n_i\) is the number of objects in the cluster \(c_i\) in
\(\pi_a\), \(e\) \(n_j\) is the number of objects in the cluster \(c_j\) in \(\pi_b\), \(k_a\) and \(k_b\) are the number of clusters in the partions \(\pi_a\) and \(\pi_b\).

\[
ARI = \frac{\sum_{k=1}^{k_a} \sum_{j=1}^{k_b} \binom{n_{ij}}{2} - \sum_{k=1}^{k_a} \binom{n_i}{2} \sum_{j=1}^{k_b} \binom{n_j}{2}}{\frac{1}{2} \sum_{k=1}^{k_a} \binom{n_i}{2} + \sum_{j=1}^{k_b} \binom{n_j}{2} - \sum_{k=1}^{k_a} \binom{n_i}{2} \sum_{j=1}^{k_b} \binom{n_j}{2}} / \binom{n}{2},
\]

(33)

At last, Luo et al. [65] modeled the similarity matrix for spectral clustering into objective functions. They assume
that \(y = Ax\) is a linear equation of an under-determined system, where \(A \in \mathbb{R}^{M \times N}\) is a full-rank and over-complete
matrix is called a over-complete dictionary, \(y \in \mathbb{R}^M\) is called a measurement vector, and \(x \in \mathbb{R}^N\) is a sparse vector. So,
they use \( x \) and \( A \) to reconstruct \( y \). For that the SParsity (SP) and Reconstruction Error (RE) should be minimized.

\[
SP = \|x\|_0, \quad \text{where } \|\cdot\|_0 \text{ counts the number of nonzero values in a vector.}
\]

\[
RE = \|Ax - A\|_2^2, \quad \text{where } \|\cdot\|_2^2 \text{ is the Euclidean norm on signals of a square matrix}
\]

\[ (34) \]

\[ (35) \]

3.4 Evolutionary Operators

The evolutionary optimization relies on the evolutionary operator to generate new solutions. Usually, it is applied one operator of crossover and one operator of mutation, considering traditional evolutionary operators or clustering designed operators.

One-Point Crossover, Two-Points Crossover, Shuffle Crossover, Uniform Crossover, SBX-Simulated Binary Crossover, Polynomial mutation, Uniform mutation are examples of traditional operators. In the One-Point Crossover, one crossover point is considered along the length of the parents chromosomes, and the genes following the crossover point in a parent are swapped with the genes in the other parent. Two-Point Crossover considers two crossover points along the length of the chromosome of each parent, such that the interval of genes between these two points is swapped. The Shuffle crossover is similar to one-point crossover, in which a single crossover position is selected, and before the variables are exchanged, they are randomly shuffled in both parents. In the Uniform Crossover, for each position of the chromosome, a random decision is made on whether the swapping of genes should be done or not. SBX uses a probability density function that simulates the One-Point Crossover in binary-coded representation. The polynomial mutation considers a polynomial probability distribution to perturb a solution. The uniform mutation replaces the value of the chosen particular slot position with a uniform random value selected considering a specified upper and lower bounds for that position.

In terms of the clustering designed operators, the representation and clustering criteria are taken into consideration. For example, the perturbation or replacement of center, centroid, or medoid is applied in the algorithms that use a prototype-based encoding to shift a randomly selected center slightly from its current position or replace the position of the cluster prototype according to a criterion; the exchange the prototypes considers two parents in which there is an exchange of centroids to generate a new solution. Also, there are operators designed to split the objects of a cluster or merge two or more clusters to generate new solutions. Handl and Knowles [41] presented the neighborhood-based mutation that is applied on the graph-based representation, replacing an existing link in the graph with another link to one of the randomly selected nearest neighbors. In [5, 11], Cheng and Church’s (CC) algorithm was adapted to be applied as a mutation operator. The CC algorithm considers three steps (multiple node deletion, single node deletion, and node addition) to iteratively perform the removal and addition of rows and columns in a data expression matrix. As a mutation operator, only row operations are performed to preserve specific data properties. Besides that, Faceli et al. [28] introduced the use of clustering ensemble (see Section 2.1) as a crossover operator, usually associated with the label-based encoding. Instead of the usual combination of all base partitions at the same time to generate a single consensus partition, pairs of partitions are combined, iteratively, in a multi-objective evolutionary process.

3.5 Obtaining a Final Solution

The Final Selection operation is applied to restrict the number of clustering solutions presented to the decision-maker or data specialist. One way to select the final set of solutions is by applying CVIs. For example, in [112], Pakhira, Bandyopadhyay and Maulik (PBM) [81], and DB indices were used to single out the optimal solution. In [69, 70], the
solution with the highest value of the $Sep_{graph}$ in the Pareto Front was considered the best solution to be selected. In [120], a new indicator called projection similarity validity index (PSVIndex) was designed to select the best solution and cluster number. In [24], the EMOC approach uses an overall ranks of nine CVIs to determine the final set of solutions: C index [3], COSEC - Compactness and Separation Measure of Clusters [88], $DB$, $Dunn$, $Dev$, $Entropy$, $XB$, Purity [97] and F-Measure [54]. In particular, [65] the nondominated solutions are used to construct a standard adjacency matrix, and the measurement Ratio Cut [118] provides a way to select a final trade-off solution.

Another way to select final solutions is by applying the knee-based approaches. The knee method presented by Handl and Knowles [41] compares the final set of solutions and a control front. The solution corresponding to the largest distance between the actual nondominated front and the control fronts is chosen to be the final solution, corresponding to the 'knee' (the point of inflection) of the nondominated front. In [20, 116], the best clustering result is defined by the 'elbow' method, which consists in picking the 'elbow' or 'knee' of the curve in the nondominated front.

Besides that, clustering ensemble methods (see Section 2.1) also are used in the selection of the final solutions, in which the nondominated solutions are used as base partitions to generate the consensual partition. In this context, the use of the ensemble methods as crossover operators (associated with the no use of a mutation operator) may lead to a reduced number of solutions in the final population because it does not produce a large number of similar partitions in the evolutionary process [28].

3.6 Multi-objective Evolutionary Clustering Evaluation

In terms of the clustering result, in general, to evaluate the general performance of the multi-objective evolutionary clustering is applied an external validity index, as ARI (Equation (33)). Besides that, the analysis of internal criteria also can be applied to investigate specific data structures. For example, in [92, 93], $H$, $Sep_{AL}$, $Dunn$, and $Dev$ are evaluated to analyze the general behavior of the EMOC approaches regarding each criterion. In [22, 23], they compare their approaches with other ones based on the $DB$, $H$ and $Sep_{AL}$.

It is important to observe that, rather than only use the CVIs to evaluate the algorithm performance, the evaluation of the optimizer can generate essential information regarding the modeled problem. They verify how well the final population reaches the goal to obtain a converging and diverse set of solutions compared to the initial population. There is a variety of quality indicators applied to measure the multi-objective optimization, as presented in [94], as like IGD - Inverted Generational Distance and HV - Hypervolume.

The IGD is computed on the objective space, which can be viewed as an approximate distance from the Pareto front to the solution set in the objective space. So, given a set of solutions $S$ and a set of $R$ uniformly distributed representative points of the PF, the IGD measure is computed according to Equation (36), where $d(r, S)$ is the minimum Euclidean distance between $r$ and the points in $S$, and $|R|$ is the cardinality of $R$. A lower IGD result refers to a better quality of $S$ [124].

$$IGD(S, R) = \sum_{r \in R} \min_{S} d(r, S) \frac{1}{|R|}$$  \hspace{1cm} (36)

The HV measures the volume of the area enclosed by the set and a reference point specified by the user. The hypervolume formula is given by Equation (37), where $vol$ refers to the Lebesgue measure and $z = \{z_1, ..., z_m\}$ is a given reference point, nadir point $z_{nad}$. A nadir point corresponds to the worst Pareto-optimal solution of each objective, and the nadir objective vector represents the worst value of each objective function corresponding to the
A Survey of Evolutionary Multi-Objective Clustering Approaches

entire Pareto-optimal set \[133\].

\[
HV(S) = \text{vol}(\bigcup_{x \in S} [f_1(x), z_1] \times \ldots \times [f_m(x), z_m])
\]  \hspace{1cm} (37)

4 OVERVIEW OF THE MULTI-OBJECTIVE CLUSTERING STUDIES

In this section, we show data analysis on a set of Multi-objective Clustering (MOC) studies. This survey considers papers related to MOC from IEEE Xplore\(^2\), ACM Digital Library\(^3\) and Scopus\(^4\). These article repositories contain the most important journal papers and conference proceedings, in the computer science and engineering domains. We chose the words "multi-objective", "multiobjective", "many-objective" as keywords related to the optimization with multiple objectives, associated with the "clustering" term to search by title the article related to multi-objective clustering. The article mapping was limited to considers English-language papers that were published before the year 2021. The search result is 231 papers from IEEE Xplore, 30 papers from ACM Digital Library, and 533 papers from Scopus, totaling 794 papers. Then, duplicated papers were removed. Finally, we have analyzed the main contents in the resulting set of documents, closing with 358 papers. In the following, we discuss statistics of publications of MOC works.

![Fig. 4. The number of publications related to MOC from 2002 to 2021.](https://example.com/image)

Fig. 4 shows the number of publications related to MOC that appeared in both journals and conferences over the years. It provides information how the MOC field is evolving, based on the amount of papers were published. The first indexed article found was published in 2002 \[135\], a conference paper in the Annals of the New York Academy of Sciences. In the same way, most of the articles published between 2002 and 2008 were published in conferences. From 2009, we observe a substantial increase in journal papers. Between 2008 and 2016, we verified a certain equilibrium of the number of articles published in conferences and journals, except in 2012, the number of conference papers had increased abnormally, without a specific explanation. Finally, in the last four years (2017-2020), the number of articles published in journals substantially incremented. In particular, in 2019, the number of publications in journals was almost three times greater than the number of papers presented in conferences. In 2020, we can notice that the total number of

\(^2\)https://ieeexplore.ieee.org/Xplore/home.jsp
\(^3\)https://dl.acm.org/
\(^4\)https://www.scopus.com/wsearch/form.uri?display=basic

Manuscript submitted to ACM
papers significantly decreased compared to 2018 and 2019. One reasonable motivation was the covid pandemic, which motivated periods of suspension of not essential activities, and some conferences worldwide were canceled or postponed.

Regarding the optimization approach, considering the general classification of the metaheuristics presented by [103], we observed that most studies applied evolutionary optimization. Fig. 5 presents the relation between the number total of articles and the evolutionary optimization articles, including memetic and hybrid approaches that include other methods associated with the evolutionary approach. In the early years, almost all MOC papers rely on the evolutionary approach. In the middle years, the use of other optimization methods was observed, as Artificial Immune system-inspired [111], Differential Evolution-based [26], Simulated Annealing-based [6], Particle swarm-based [89]. In the mapped articles, the first occurrence of these approaches was between 2007 and 2009. In the last years, the use of a variety of other optimization methods was also verified, as other nature-inspired algorithms [103], among others. In Section 6 we present some algorithms considering the most relevant metaheuristics.

![Fig. 5. Relation between total articles and evolutionary-based optimization articles.](image)

At last, we also verified the main topics considered in these almost twenty years of research in MOC. Fig. 6 presents the word cloud of the keywords and the indexed terms in MOC papers. Most of the terms refer to the optimization methods and application field of the MOC. The same meaning terms (single/plural and case forms of the terms) in the cloud are filtered, so "genetic algorithm", "Genetic Algorithm", and "genetic algorithms" are all treated as "Genetic algorithm". Larger words in the word cloud are the ones used more frequently in the papers analyzed.

From this word cloud, although keywords related to the algorithm such as "Clustering algorithms", "multi-objective optimization" are understandably used more frequently, it is notable that the keywords "multi-objective optimization" and "Genetic algorithm" has been attracting researchers attention as a problem domain of MOC. We also observe two main application fields: Image Segmentation and Gene/Micro-Array Analysis. The second one considers words like "Biological Cell" and "Gene Expression". Other listed application fields are Document Clustering, Community detection, Software module clustering, among others.
5 EVOLUTIONARY MULTI-OBJECTIVE CLUSTERING ALGORITHMS

In this section, we present Multi-objective Evolutionary Clustering Algorithms. The most relevant EMOC works were selected considering two general indices: h-index and Scopus percentile. We filtered the articles by h-index greater than 10 to filter the conference papers and scopus-percentile greater than 50% to obtain the list of the most relevant journal papers. These values were selected to cover the A rank paper in the CORE - Computing Research and Education Association of Australasia and Qualis (a Brazilian official system to classify scientific production). These algorithms were grouped considering some common aspects, which highlights the main features or application of these approaches. The general concepts and methods applied in these EMOC approaches were introduced in Section 3: solution representation and initialization methods in sub-section 3.1, MOEAs/MaOEAs in sub-section 3.2, objective functions in sub-section 3.3, final selection methods in sub-section 3.5.

5.1 EMOC approaches for General Purpose

5.1.1 MOCK-based works. One of the most popular algorithms is MOCK - Multi-Objective Clustering with automatic k-determination [40–42]. The MOCK algorithm uses locus-based adjacency graph representation of partitioning, initialization with MST-clustering and KM methods, two objective functions: \( D_{ev} \) and \( C_{on} \). The PESA-II was the MOEA applied in this approach. The adjacency graph representation promoted the use of specific genetic operators for the clustering problem, like the neighborhood-based mutation operator and uniform crossover. After the optimization process and the generation of final clustering solutions, MOCK uses an automatic \( k \)-determination scheme to choose the best clustering solution from a set of solutions with a knee-based strategy.

MOCK was studied along with other works, as follows. Matake et al.[68] provided an approach (MOCK-Scalable) to improve the final selection of solutions in large-scale data, based on a scaling filter to reduce the solutions in the front. Tsai et al. [112] proposed the MIE-MOCK - Multiple Information Exchange Multi-Objective Clustering with automatic \( k \)-determination. MIE-MOCK algorithm uses pool of crossover and mutation operators with random selection, and provided a final selection of solutions based on two CVIs: PMB and \( D_{B} \). In [43], Handl and Knowles analyzed four pairs of objective functions for multi-objective clustering, including an analysis of MOCK objective functions.
In [44], Handl and Knowles analyzed the use of evidence accumulation to support the post-processing of the clustering solutions returned by the MOCK. In [34, 35], Garza-Fabre et al. proposed the Δ-MOCK, providing a new encoding to improve the MOCK scalability and other specific modifications to improve the convergence of the solutions. Zhu et al. [131] provided the Δ-EMaOC - Evolutionary Many-Objective Optimization Clustering, improving the general architecture of the Δ-MOCK to optimize five objective functions. Δ-EMaOC algorithm uses MaOEAs (SPEA-II-SDE [60], NSGA-III [121], MOEA/DD [59] and RVEA [13]) instead of an MOEA (NSGA-II). In general, these approaches are applied to detect clusters in heterogeneous structured data, considering a continuous data type and a crisp clustering.

5.1.2 EMOC for Categorical Data. In particular, some EMOC approaches were designed for categorical data clustering, where the data objects are defined over categorical attributes (instead of using the continuous data type that is applied in most of the other approaches). For example, Handl and Knowles [41] presented the MOCK-medoid, a MOCK extension for multi-objective clustering around medoids for categorical data. Mukhopadhyay and Maulik [72], like [41], introduced a medoid-based EMOC, the MOGA-medoid, to deal with categorical data. MOGA-medoid algorithm uses the NSGA-II to optimize the $\text{Sil}$ and $\text{Dee}$ (computed in terms of the medoids, instead of the centroids), applying the one-point crossover and a medoid-based replacement mutation designed to consider a center-based solution encoding. In this context, Dutta et al. [22, 23] provided a specific MOEA, the Hybrid MOGA, to optimize $H$ and $\text{Sep}_{AL}$. The main contribution of this work relies on the use of this new MOEA with the Pairwise Crossover [32], the replacement (substitution) Mutation, and the local searching power of K-modes (or KM) to deal with continuous and categorical features in the dataset.

In this context, Mukhopadhyay et al. [75] provided a multi-objective genetic fuzzy clustering of categorical attributes (MOGA-fuzzy), in which is applied a uniform crossover center-based replacement mutation, considering the NSGA-II to optimize the $J_m$ and $\text{Fuzzy}$. They applied a specific selection method to generate the final solution, in which the points assigned to the same cluster by at least 50% of the clustering solutions are taken as the training set, and the remaining points are assigned a class label using k-nearest neighbor ($k$-nn) classification. In [76], they provide a new version of the MOGA-fuzzy (MOGA-fuzzy2), in which another combination of the evolutionary operator (One-Point/Mode replacement) was applied. In both versions of the MOGA-fuzzy, they use ensemble-based selection to obtain the final set of solutions with distinct ensemble methods.

Zhu and Xu [130] introduced the MaOFcentroids, a many-objective fuzzy centroids clustering algorithm for categorical data. MaOFcentroids algorithm uses fuzzy membership matrix encoding (a matrix with the degree of membership of each object), and the NSGA-III with adapted operators that consider the number of the clusters and the membership of the solutions. It simultaneously optimizes five CVIs ($\text{CDCS}$, $DB$, $CH$, $CCI$, and $XB$) and uses a specific clustering ensemble for categorical data, the SIVID - Sum of Internal Validity Indices with Diversity[127] to the final selection of solutions.

The most recent work of Dutta et al. [24] introduce the MOGA-KP, an approach with automatic k-determination applied to deal with different types of features (continuous and categorical and missing feature values). It is an extension of the previous works [22, 23], in which is analyzed other aspects of the evolutionary process, as the use of other evolutionary operators, improvements in the local search operators. MOGA-KP algorithm uses a ranking of nine CVIs ($C$ index, $COSEC$, $DB$, $Dunn$, $Dev$, $Entropy$, $XB$, $Purity$, $F$-Measure) to determine the final set of solutions.

5.1.3 EMOC for Bi-Clustering. One specific line of studies is the Bi-clustering, which consists of simultaneous partitioning of the set of samples and the set of their attributes into subsets (classes), the goal is to find one or all (possibly overlapping) sub-matrices of a given matrix, each of which shares a pre-defined property over the elements across all its columns (or rows). Each such sub-matrix is called a bi-cluster. Bousselmi et al. [11] presented the BI-MOCK,
that extends the solution encoding of MOCK to the case of bi-clustering by adding a subset of columns (conditions) to each chromosome. BI-MOCK algorithm uses the Two-points crossover adapted for variable-size chromosomes and the CC algorithm as mutation operator to optimize an $Var$, and the size of the bi-cluster in the PESA-II. Bechikh et al. [5] presented the MOBICK - Multi-Objective BI-Clustering with automated $k$ deduction, that extends Bousselmi et al. [11] work. MOBICK algorithm uses the $\Delta$-MOCK reduced encoding, the uniform crossover adapted for bi-clustering conditions, and the CC algorithm as mutation operator to also optimize an $Var$, and the size of the bi-cluster in the PESA-II.

5.1.4 EMOC for Subspace Clustering. Another line of studies considers the Subspace Clustering, that is an extension of traditional clustering that seeks to find clusters in different subspaces within a dataset. Zhu et al. [129] introduced the MOSSC - Multi-Objective evolutionary algorithm-based Soft Subspace Clustering, that optimize the $SSBX$ and $J_{wem}$ in the NSGA-II, considering a center and weight encoding to avoid trapping in local minima and thus obtain more stable clustering results. Xia et al. [120] presented the MOEASSC - Multi-Objective Evolutionary Approach-based Soft Subspace Clustering, which also uses a mixed encoding (center and weight-based). MOEASSC differs from the MOSSC in terms of the pair of objectives ($J_{wem}$ and $J_{Add}$), and the use of a local search operator based on the KM. Zhou and Zhu [122] introduced the MOKCW - Multi-Objective Kernel Clustering algorithm with automatic attribute Weighting that extends MOSSC and MOEASSC. MOKCW uses the MOSSC objective functions adapted to consider kernel distance. They also improved the final selection method of the MOEASSC by applying a clustering ensemble method (MCLA and HBGF) associated with the PSVIndex.

5.1.5 Ensemble-based EMOC. Another specific approach was proposed by Faceli et al. [28], the MOCLE - Multi-Objective Clustering Ensemble. The main idea around this approach is the use of the clustering ensemble methods (Section 2.1) as crossovers operators to combine partitions and extract agreed patterns to generate new solutions in the evolutionary optimization process. MOCLE is a framework that uses a label-based representation; the initial population is generated with various clustering methods to detect different clusters formats, like SL, AL, KM, SNN. The original implementation of the MOCLE [28] provide two MOEAs: NSGA-II and SPEA-II, to optimize the $Deo$ and $Con$; two crossovers operators: MCLA - Meta Clustering Algorithm [109] and HBGF - Hybrid Bipartite Graph Formulation [30]; however it does not uses any mutation operator.

This general idea of using clustering ensemble methods as crossover operators also was applied in other works. For example, Faceli et al. [29] introduced the MOCLE in the context of gene expression datasets applying an additional objective, $ConP$ (the connectivity index based on the Pearson Correlation) and a new set of clustering methods to generate the initial population (AL, CL, KM, and SPC). Liu et al. [63] introduced the IMOCLE - Improvement of the Multi-Objective Clustering Ensemble algorithm, in which $Sim$ was added to the three objective functions defined by Faceli et al. [29]. In general, these approaches also are applied to detect clusters in heterogeneous structured data, considering a continuous data type and crisp clustering.

5.1.6 Fuzzy Clustering methods-based EMOC. Another line of works considers the integration of the general concepts of the existing fuzzy clustering algorithms, as FCM and FRC - Fuzzy Relational Clustering, with a multi-objective evolutionary approach (NSGA-II). Di Nuovo et al. [18], Wikaisuksakul [119] and Dong et al. [19] presented fuzzy approaches integrating the NSGA-II with the FCM [7]. Di Nuovo et al.[18] introduced the NSGA-II&FCM that optimize the number of features and $XB$ index to discover the best number of groups while pruning the features to reduce the dimensionality of the dataset. NSGA-II&FCM algorithm uses a specific solution encoding that considers the FCM.
parameters (number of the cluster \(k\) and FCM fuzzifier \(m\)) and the features weights. Wikaisuksakul [119] presented the FCM-NSGA that optimize the \(J_m\) and \(\text{Sep}_{\text{fuzzy}}\), considering SBX and polynomial operators. Dong et al. [19] introduced the ADNSGA2-FCM that optimize the \(DB\) and \(J\) indexes. ADNSGA2-FCM uses two solutions encoding, center-based and fuzzy membership matrix (a matrix with the degree of membership of each object), and the uniform mutation with two new crossover operators, Nearest Neighbor Matching Crossover Operation (an exchange of center in the nearest neighbor to produce solutions with the same number of clusters) and Truncation and Stitching Crossover Operation (an exchange of a set of center positions are performed to produce solutions with a different number of clusters). Moreover, they introduced an adaptive mechanism applied to compute the crossover and mutation probabilities that are changed according to the fitness of the current population. On the other hand, Paul and Shill [82] propose the FRC-NSGA/IFRC-NSGA, a hybrid method that combines the FRC algorithm [104] and the NSGA-II to optimize the \(J_m\) and \(\text{Sep}_{\text{fuzzy}}\).

### 5.1.7 Spectral Clustering-based EMOC

Some works consider the Spectral-Clustering as a base to design the EMOC approaches. MOGGC - Multi-Objective Genetic Graph-based Clustering Algorithm [69] provided a new objective function pair, the separation of clusters (\(\text{Sep}_{\text{Graph}}\)) and a graph continuity metric (DCD), to achieve lower memory consumption and increased the clustering solutions quality. MOGGC was extended by the CEMOG - CoEvolutionary Multi-Objective Genetic Graph-based Clustering [70], a partitional \(k\)-adaptive spectral clustering algorithm that uses a strategy based on island-model and a graph topology to migrate individuals from sub-populations. It does not require inputting the initial number of clusters required in the MOGGC. In this context, Luo et al. [65] introduced the framework SRMOSC, which uses a sparse representation with a specific pair of operators that considers the sparsity properties. SRMOSC uses the \(\text{SP}\) and \(\text{RE}\) as objective functions to be optimized in the NSGA-II or MOEA/DD.

### 5.1.8 Multiple Distance Measures-based EMOC

Other approaches consider the use of different distance functions in the objective functions. Liu et al. [61] introduced the MOECDM - Multi-objective Evolutionary Clustering Based on Combining Multiple Distance Measures and the MOEACDM - Multi-objective Evolutionary Automatic Clustering based on Combining Multiple Distance Measures, that considers a CVI computed with distinct distance functions to define distinct objective functions. In both approaches, they use a label-based encoding and an NCUT pre-clustering[101] in the initialization, but in the MOECDM, part of the individuals are generated by random. They also adapted the crossover and mutation operators, in which the probabilities are adjusted along with the generations. MOECDM was designed to detect the desirable cluster number automatically, using the \(\text{Sep}_{CL}\) index computed with Euclidean distance and Path distance as objectives functions. MOEACDM was designed to detect compact clusters, using the \(\text{Modularity}\) also computed with Euclidean distance and Path distance as objective functions.

### 5.1.9 Multi-k-clustering-based EMOC

Other approaches consider the multi-\(k\)-clustering with a posteriori method, where \(k\) is taken as an objective function, differing from the automatic data clustering methods, as MOCK, that considers \(k\) an inner aspect in the decision variable, obtained by the optimization of clustering criteria.

For that, Du et al.[20] introduced a specific solution representation, the linked-list based encoding. They use the fellowship between the objects instead of the label-based relationship to define the clusters, in which each cluster has all their elements linked, similar to the relation of the nodes presented by Handl and Knowles [41]. This representation was applied in the MOGA-LL[20], an EMOC approach that optimizes the \(TWCV\) and \(k\) as objective functions in the NPGA, considering two particular operators: (i) an adapted one-point crossover that allows different clusters to exchange
partial contents and may split a cluster into two; (ii) link-replacement mutation, in which a sub-group of objects is associated to another cluster instead of just a different node.

Wang et al. [116] proposed the EMO-k to demonstrate the importance of the conflict between the objective functions. They showed evidence that $Var$ and $k$ are not always conflicting between two individuals and promoted a transformation of the variance ($Var$) formulation, $(1 - \exp^{-1.\text{Var}}) - k$, to solve this problem. In [114], this same pair of objective functions were explored considering a new MOEA based on constrained decomposition with grids (CCDG-K). Both EMO-K and CCDG-K define the best clustering result (the optimal k) by the “elbow” method [38].

5.1.10 Specific MOEA for EMOC. As previously presented, Dutta et al. [22, 23] provided a specific MOEA, the Hybrid MOGA designed for categorical data. Besides that, other particular approach is the VRJGGA - Variable-length Real Jumping Genes Genetic Algorithm introduced by Ripon et al. [92]. The VRJGGA is an EMOC algorithm that extends Jumping Genes Genetic Algorithm (JGGA) [67] and the survival selection method of the NSGA-II. The JGGA considers jumping genes operations before the evolutionary operators to improve the diversity of solutions. The VRJGGA uses a centroid-based encoding associated with the modulo crossover [108] (an adapted one-point crossover, where each child is a set of completely specified subsolutions) and the polynomial mutation, to optimize the Entropy and Sep$_A$. In [93], they provided new features to VRJGGA, introducing two local search methods, a probabilistic cluster merging, and splitting for the clustering improvement. Ripon and Siddique [91] also applied the extended version of the JGGA to EMOC, introducing the EMCOC - Evolutionary Multi-objective Clustering for detecting overlapping clusters. EMCOC introduces a new chromosome representation and cluster-assignment method, in which each data point is a candidate center, and a binary encoding is applied to define whether a data point is a center or not.

5.1.11 Other MOC approaches. Some works consider other objective functions and provide other features to the EMOC approaches. For example, Kirkland et al. [53] presented the Multi-Objective Clustering algorithm (MOCA), that optimizes three objective functions, $AWGSS$, $ABGSS$, and $Con$ in the NSGA-II. Sert et al. [98, 99] presented the MOC-HCM, that uses five objective functions: $K_{\text{mid}}$, $K_{\text{med}}$, $K_{\text{med}}$, $K_{\text{med}}$ and $EWCD$ in the NSGA. The MOC-HCM algorithm uses a binary representation, a local search operator (k-mode-based operator) that reassigns the instances to the closest clusters in terms of their frequencies and a new final selection method based on a new metric, the H-Confidence Metric (HCM).

In Table 1, we summarize the main features of each approach, considering the method applied in the initialization, the multi-objective algorithm, the objective functions, evolutionary operators, and selection methods. This table was formulated considering the publishing chronology. We used acronyms for some words: Ad. for Adapted, Repl. for Replacement, and Mod. for Modified.

Besides the above-mentioned works, we also found specific approaches, in which their main features consider some particular methods, as follows. Özyer and Alhaj [80] applied the divide and conquer approach in an iterative way to handle the clustering process and improve the performance of the evolutionary algorithm. Zheng et al. [128] extended algebraic operations of gene expression to propose a multi-objective gene expression programming for clustering. Garcia-Piquer [33], focused on reducing the impact of the volume of data in the EA by means of the stratification of the complete data set into disjoint strata and alternate them in each cycle of the GA. Liu et al. [62] improved the performance of multi-objective soft subspace clustering algorithms for clustering high-dimensional data by a transfer learning-assisted multi-objective evolutionary clustering framework with MOEA/D.
| Year  | Article               | Representation               | Initialization | MOEA/MaOEA | Objectives          | Crossover/Mutation                      | Final Selection  |
|-------|-----------------------|-----------------------------|----------------|------------|---------------------|-----------------------------------------|-----------------|
| 2005  | MOCK [40–42]          | Adjacency graph-based       | MST, KM        | PESA-II    | Con and Dev         | Uniform/ Neighborhood-based             | Knee-based      |
| 2005  | MOCK-medoid [41]      | Adjacency graph-based       | KM             | PESA-II    | Var and Dev         | Uniform/ Neighborhood-based             | Knee-based      |
| 2005  | MOGA-LL [20]          | Linked-list-based           | Random         | NSGA-II-based JGGA Extension | TWCV and k                                 | Ad. One-Point/ Link Repl.                |                 |
| 2006  | VRJGGA [92, 93]       | Centroid-based              | Random         | NSGA-II    | Entropy and Sep_AL  | Modulo/ Polynomial                      |                 |
| 2006  | MOCLE [28]            | Label-based                 | KM, AL, SL, SNN | SPEA2 or NSGA-II | Con and Dev         | HBGF or MCLA/—                          |                 |
| 2006  | MOGA-medoid [72]      | Medoid-based                | Random         | NSGA-II    | Dev and Sil         | One-point/ Medoid Repl.                 |                 |
| 2006  | MOCK-scalable [68]    | Adjacency graph-based       | MST, KM        | SPEA2      | Con and Dev         | Uniform/ Neighborhood-based             | Knee-based      |
| 2006  | MOGA-fuzzy [75]       | Mode-based                  | Random         | NSGA-II    | Jm and Sep_fuzzy    | Uniform / Mode Repl.                     |                 |
| 2006  | NSGA-II & FCM [18]    | FCM parameters and features weights | FCM           | NSGA-II    | Number of features and XB   | SBX / Polynomial                       |                 |
| 2006  | MOGA-fuzzy2 [76]      | Mode-based                  | Random         | NSGA-II    | Jm and Sep_fuzzy    | One-Point / Mode Repl.                  | Ensemble-based (Majority vote) |
| 2006  | EMOCOC [91]           | Binary center-based         | Random         | NSGA-II-based JGGA Extension | Entropy and Sep_AL                  | Center exchange / Not specified         |                 |
| 2011  | MOCA [53]             | Medoid-based                | Random         | NSGA-II    | AWGSS, ABGSS and Con | Centroids exchange/ Split, Merge, Centroid Repl. |                 |
| 2011  | MOC-HCM [98, 99]      | Binary-based                | Not specified  | NSGA       | Kmid, Kmed, Kmwd, Kmwd and EWCD | Shuffle / Not specified | Ensemble-based (H-confidence) |
| 2012  | IMOCLE [63]           | Label-based                 | KM, AL, CL, SPC | NSGA-II    | Con, ConP, Dev and Sim | MCLA/—                                |                 |
| 2012  | Hybrid MOGA [22, 23]  | Centroid-based              | Random         | Specific MOEA | Sep_fCL and H     | Pairwise/Centroid Repl.                  |                 |
| 2012  | MIE-MOCK [112]        | Adjacency graph-based       | MST, KM        | PESA-II    | Con and Dev         | Uniform, One-Point and Two-Point / Neighbourhood-based, Split and Merge |     |
| 2012  | MOSSC [129]           | Center and weight-based     | Random         | NSGA-II    | SXXB and f_sum      | SBX/ Polynomial                        | Ensemble-based (HBGF) |
| 2013  | MOGCC [69]            | Label-based                 | Random         | SPEA2      | sep_graph and DCD   | Labels Exchange/ Adaptive              | sep_graph       |
| 2013  | MOEASSC [120]         | Center and weight-based     | Random         | NSGA-II    | J_sum and J_add     | One Point/Uniform                           | PSVIndex         |
| 2014  | CEMOG [70]            | Label-based                 | Random         | SPEA2      | sep_graph and DCD   | Labels Exchange/ Adaptive              | sep_graph       |
| 2014  | FCM-NSGA [119]        | Center-based                | Random         | NSGA-II    | Jm and Sep_fuzzy    | SBX/ Polynomial                        |                 |
| 2016  | SRMOSC [65]           | Sparse-based                | Neighbor-based | NSGA-II or MOEA/DD | RE and SP | Specific operators based on the sparsity property | Ratio cut-based |
| Year | Articles                     | Representation                          | Initialization | MOEA/MaOEA   | Objectives                   | Crossover/Mutation | Final Selection |
|------|------------------------------|----------------------------------------|----------------|--------------|------------------------------|--------------------|-----------------|
| 2017 | Δ-MOCK [34, 35]              | Reduced Adjacency graph-based          | MST            | NSGA-II      | Con and Var                  | Uniform/ Neighborhood-based | –               |
| 2017 | BI-MOCK [11]                | Reduced Adjacency graph-based with     | MST, KM        | PESA-II      | Var and Size of the bi-cluster | Ad. Two-Points/ CC | –               |
| 2018 | MOKCW [122]                 | Center and weight-based                 | Random         | NSGA-II      | Ad. J_{con} and Ad. SSBX     | One-Point/ Uniform  | PSVIndex and Cluster ensemble (HBGF or MCLA) |
| 2018 | EMO-k-clustering [116]       | Centroid-based                          | Random         | NSGA-II      | Mod. Var and k               | SBX/ Polynomial    | Elbow-based     |
| 2018 | FRC-NSGA/ IFRC-NSGA [82]    | Center-based                            | Random         | NSGA-II      | J_{m} and Sep_{NSFuzzy}      | SBX/ Polynomial    | –               |
| 2018 | Δ-EMaOC [131]               | Reduced Adjacency graph-based with      | MST            | NSGA-II/ RVEA/ MOEA/DD/ SPEA-II-SDE | Con, Var, Dunn, DB and CH | Uniform/ Neighborhood-based | –               |
| 2018 | MOECDM [61]                 | Label-based                             | Random and NCUT | NSGA-II      | Sep_{CL} with two different distances | Ad. Uniform/ Ad. Uniform | –               |
| 2018 | MOEACDM [61]                | Label-based                             | NCUT           | NSGA-II      | Modularity with two different distances | Ad. Uniform/ Ad. Uniform | –               |
| 2018 | ADNSGA2-FCM [19]            | Center and fuzzy membership matrix-based | Random         | NSGA-II      | DB and I                     | Neighborhood-based, Truncation and Stitching / Uniform | Ensemble-based (Majority vote) |
| 2018 | MaOfFcentroids [130]        | Fuzzy membership matrix-based           | Random         | NSGA-III     | CDCS, DB, CH, CC1, and XB   | Uniform/Specific Mutation for Membership Repl. | Ensemble-based (SIVID) |
| 2019 | MOBICK [5]                  | Reduced Adjacency graph-based with      | MST, KM        | PESA-II      | Var and Size of the bi-cluster | Ad. Uniform/ CC | –               |
| 2019 | MOGA-KP [24]                | Centroid-based                          | Random         | Specific MOEA | Sep_{CL} and H               | One-Point/ Polynomial | –               |
5.2 EMOC Approaches Designed for Specific Applications

In this section, we present approaches designed for specific applications. Each algorithm considers the particularities of problem application to define the representation of the solutions, the objective functions or/and the evolutionary operators. It promotes a generation of a variety of configurations, so we limit to list some algorithms designed for each following application.

5.2.1 Association rule learning. Association rule learning is a rule-based machine learning method for discovering interesting relations between variables in large databases. Alhajj and Kaya [1, 51] provided an EMOC approach for fuzzy association rules mining to automatically cluster values of a given quantitative attribute to obtain a high number of large itemsets in low duration (time).

5.2.2 Document clustering. Document clustering is a data/text mining technique that makes use of text clustering to divide documents according to various topics. Lee et al. [55] proposed a method of enhancing Multi-Objective Genetic Algorithms for document clustering with parallel programming. Wahid et al. [113] presented a new approach for Document Clustering based on SPEA-II, that explores the concept of multiple views to generate multiple clustering solutions with diversity.

5.2.3 Gene/micro-array analysis. The Gene/Micro-array clustering analysis is applied for discovering groups of correlated genes potentially co-regulated or associated with the disease or conditions under investigation. Romero-Zaliz et al. [95] provided an EMOC to identify conceptual models in structured datasets that can explain and predict phenotypes in the immuno inflammatory response problem, similar to those provided by gene expression or other genetic markers. Li et al. [58] provided a new ensemble operators to improve the data clustering in gene expression datasets in IMOCLE [63]. Mukhopadhyay et al. [77] provide an approach that simultaneously selects relevant genes and clusters the input dataset. Mukhopadhyay et al. [78] presented an interactive approach to multi-objective clustering of gene expression patterns considering an adapted NSGA-II, in which inputs from the human decision-maker (DM) are taken to learn which objective functions are more suitable for the datasets. [25] presented an EMOC approach to identify gene clusters from a given expression dataset; in which apart from utilizing the gene expression values of the individual genes, the corresponding protein-protein interaction scores are also used while clustering the set of genes.

5.2.4 Image Segmentation. Image Segmentation consists of the process by which a digital image is partitioned into various subgroups (multiple parts or regions), often based on the characteristics of the pixels in the image. Qian et al. [87] presented a multi-objective evolutionary ensemble algorithm to perform texture image segmentation. Shirakawa and Nagao [102] introduced a variation of the MOCK [42] improving its general features to its application in image segmentation. Zhang et al. [123] provided a multi-objective evolutionary fuzzy clustering for image segmentation, considering the original FCM energy function to preserve image details and the function based on local information to restrain noise minimized as objective functions to be minimized by MOEA/D. Zhao et al. [125, 126] introduced the use of intuitionistic fuzzy set (IFS) concept and multiple spatial information to generate an EMOC approach to overcome the effect of noise in image segmentation.

5.2.5 Software module clustering. Software module clustering refers to the problem of automatically organizing software units into modules to improve program structure. Praditwong et al. [86] provided a multi-objective formulation of the software module clustering problem considering a two-archive Pareto optimal genetic algorithm. Barros [4] provided an analysis of the effects of composite objectives in multi-objective software module clustering.
5.2.6 **Network community detection**. Network community detection refers to the procedure of identifying groups of interacting vertices in a network depending upon their structural properties to unveil the dynamic behaviors of networks. Folino and Pizzuti [31] provided an approach for the detection of communities with temporal smoothness formulated as an EMOC. Hariz et al. [2] reformulate the community detection problem as an EMOC model that can simultaneously capture the intra and inter-community structures based on functions inspired from different types of node neighborhood relations. Shang et al. [100] introduced an EMOC approach based on $k$-nodes update policy and similarity matrix for mining communities in social networks. [85] provided a framework for detecting community structure in attributed networks, introducing a post-processing local search procedure that identifies those communities that can be merged to provide higher quality community divisions.

5.2.7 **Web recommendation**. Web topic mining and web recommendation consider the problem of extracting web navigation patterns, based on the interest of a user, to be applied in the recommender systems to guide the users during their visit to a Web site. Demir et al. [17] presented EMOC approaches to cluster Web user sessions in a Web page recommender system. Morik et al. [71] investigated the problem of finding alternative high-quality structures for (Web) navigation in a large collection of high-dimensional data, and they provided a formulation of FTS (Frequent Terms Set) clustering as a multi-objective optimization problem.

5.2.8 **WSN - Wireless Sensor Network topology management**. There are several challenges in designing WSN because the sensor nodes have limited resources of energy, processing power, and memory. In this context, the clustering technique can organize nodes into a set of groups based on a set of pre-defined criteria to improve their usage. Peiravi et al. [83] provided an EMOC approach whose goal was to obtain clustering schemes in which the network lifetime was optimized for different delay values. Hacioglu et al. [37] presented an EMOC approach that can extend network lifetime while enabling high coverage and data.

5.2.9 **Other applications**. Wang et al. [117] proposed an approach to solve the circuit clustering problem in field-programmable gate array (FPGA) computer-aided design (CAD) flow. Bandyopadhyay et al. [73] introduced a multi-objective genetic clustering for pixel classification in remote sensing imagery. Wang et al. [115] and Li et al. [57] provided a multi-objective fuzzy clustering approach for change detection in Synthetic aperture radar (SAR) images. Liu et al. [64] presented an approach to automatic clustering of shapes considering a multi-objective optimization with decomposition and improvement in the shape descriptor and diffusion process (that was applied to transform the similarity distance matrix among total shapes of a dataset into a weighted graph).

### 6 OTHER METAHEURISTIC ALGORITHMS FOR MULTI-OBJECTIVE CLUSTERING

The studies of other metaheuristics in the MOC problem have expanded over the years. The top metaheuristics with the highest number of publications, besides the evolutionary approaches (194 articles), are PSO - Particle Swarm Optimization (35 articles), DE - Differential Evolution (23 articles), SA - Simulated annealing (19 articles), AIS - Artificial Immune Systems (11 articles). PSO is a stochastic optimization inspired by the social behavior of bird flocking and schooling in nature that searches the optimal region of optimization space through the interaction of particles (candidate solution) [89]. Multi-objective clustering based on the PSO presents a large number of application fields in the same way as the evolutionary approaches. DE is a stochastic direct search method approach applied for solving numerical continuous optimization problems [26]. In the mapped papers, most MOC approaches that consider DE optimization is designed for general purpose, and the specific purposed designed DE-based approaches lie on bioinformatics studies.
(micro-array analysis and disease stratification) and image clustering in remote sensing images. SA is a stochastic approach that simulates the statistical process of growing crystals using the annealing process to reach its absolute (global) minimum internal energy configuration. The MOC approach introduced by Saha and Bandyopadhyay [96] is the most popular SA-based MOC for general purposes. However, most mapped SA-based MOC approaches are designed for bioinformatics, such as micro-array analysis and patient diagnosis applications. AIS are adaptive systems inspired by the information processing mechanism of the biological immune system that has characteristics of self-organization, learning, memory, adaptation, robustness, and scalability [111]. In the mapped papers, most of the AIS-based MOC approaches analyze different aspects of these characteristics, being widely applied to SAR image segmentation.

7 DISCUSSION AND CONCLUSION

This article presented an overview of the publications on multi-objective clustering, considering the indexed papers in ACM Digital Library, IEEE Xplore, and Scopus. We selected the papers cited in this review based on metrics of the impact and relevance of the conference/journal, promoting a not-biased selection of papers and better coverage of the EMOC studies. The result was summarized in Table 1, which presents the main features of each approach, considering the general architecture of an evolutionary MOC algorithm shown in Fig. 2.

This mapping of EMOC approaches allows us to observe some patterns and obtain some insights regarding the evolutionary multi-objective clustering algorithms. For example, the choice of the objectives functions is one of the most critical factors of the optimization process. We can observe a diverse combination of clustering criteria as objective functions to the EMOC problem over the years. A common practice considers at least one compactness-based criterion associated with a connectedness-based criterion to clustering heterogeneous structured data. In the case of the centered-based clustering optimization, it is common to see other schemes of the objectives: (i) the optimization of a compactness-based criterion and the number of the clusters, (ii) a combination of the two compactness-based criteria, (iii) the optimization of a compactness-based criterion and a spatial separation-based criterion. In this last case, these different configurations of objective functions are mostly related to specific classes of clustering studies, as bi-clustering (i), categorical data clustering (ii and iii).

In general, there is not a consensus around the ideal number and the best combination of objective functions between the researchers because of the difficulty in defining appropriate clustering criteria. In this context, besides the cluster proprieties, it is essential to be wary of the conflict required of the objective functions in the choice of the objective functions to generate a diverse and convergence set of solutions. In particular, Wang et al. [116] reported that the conflicting relationship between $Var$ and $k$ is not guaranteed, where the solutions for some $k$ values are dominated during the search, and in the worst case, no solution can be found for those $k$. Also, it opens a doubt regarding the other combinations of objectives that uses the $Var$ associated with another criterion, as $Con$, where $k$ is an implicit conflicting aspect. In this way, more studies on the objective functions are required to improve the composition of objective functions and provide more information on the limitation of the existing ones.

In terms of an evolutionary multi-objective approach, we can note the use of the NSGA-II as MOEA and the recent usage of MaOEAs in two works [130, 131] considering five objective functions. In contrast, other works [63, 98, 99] considered the optimization of more than three objective functions in MOEAs (NSGA/NSGA-II). In this context, another research direction is to analyze the behavior of the other MOEAs/MaOEAs, or even other categories of multi-objective methods (see Section 3), in the clustering problem. For example, the use of diversity-based MOEAs/MaOEAs in approaches that seek for more diversity, as Liu et al. [63] - that applied a fourth objective function, $Sim$, with other three objective functions ($Con$, $ConP$, $Deo$) to obtain more diversity of solutions in the population. In particular, the
evolutionary multi-objective clustering ensemble approaches [28, 29, 58, 63], which considers clustering ensemble as a crossover operator to combine information of two or more solutions, have an inner property to reduce the number of solutions while producing new solutions. So, in general, it does not require a post-processing method to select the final set of solutions. In this context, two aspects are not fully clear: the first one is the impact of the improvement of the diversity of solutions in the population, in terms of the number of the solutions presented to the decision-maker; the second aspect is the impact of reducing the number of individuals in the evolutionary optimization because the reported results only consider CVIs, not providing an evaluation in terms of the optimization.

Regarding the final selection, we note that some approaches do not provide a final selection method, providing only an evaluation regarding the clustering process in comparison to other approaches. So, the decision-maker has to use another tool to select the best solutions in these approaches. The choice of the mechanisms to select the best solution or set of solutions also is a challenge that requires more studies. In general, the existing selection methods are the ensemble-based that provides the best solution (consensual partition); the knee-based that provides the best $k$-solution; the CVIs-based, which considers specific criteria (as ranking) to define the best set of solutions. By using the selection based on CVIs, in the same way as the objective functions, it is required to observe the limitation of the clustering criteria.

Another concern in this field is regarding the real applications and large-scale clustering problems. Some works, as [33, 35, 130] improve the scalability on designing more efficient multi-objective evolutionary algorithms; however, most of the existing multi-objective evolutionary algorithms are not well scalable to real-life applications that generate a huge amount of data. According to Mukhopadhyay et al. [79] it is a challenge for researchers to devise fast, scalable algorithms for multi-objective clustering.

At last, this paper also presented some applications of EMOC and the most relevant related papers that can be useful to researchers that are exploring EMOC for a specific purpose. Furthermore, we list other metaheuristics that were explored in studies over the years. Our mapping did not find surveys or reviews that fully cover MOC approaches with these metaheuristics, and this aspect can be explored in future works.

ACKNOWLEDGMENTS

The authors would like to thank the financial support of the National Council for Scientific and Technological Development (CNPq).

REFERENCES

[1] R. Alhajj and M. Kaya. 2008. Multi-objective genetic algorithms based automated clustering for fuzzy association rules mining. Journal of Intelligent Information Systems 31, 3 (2008), 243–264. https://doi.org/10.1007/s10844-007-0044-1
[2] B. A. Attea, W. A. Hariz, and M. F. Abdulhalim. 2016. Improving the performance of evolutionary multi-objective co-clustering models for community detection in complex social networks. Swarm and Evolutionary Computation 26 (2016), 137–156. https://doi.org/10.1016/j.swevo.2015.09.003
[3] F. B. Baker and La. J. Hubert. 1976. A graph-theoretic approach to goodness-of-fit in complete-link hierarchical clustering. J. Amer. Statist. Assoc. 71, 356 (1976), 870–878.
[4] M. O. Barros. 2012. An Analysis of the Effects of Composite Objectives in Multiobjective Software Module Clustering. In Proceedings of the 14th Annual Conference on Genetic and Evolutionary Computation (Philadelphia, Pennsylvania, USA) (GECCO ’12). Association for Computing Machinery, New York, NY, USA, 1205–1212. https://doi.org/10.1145/2330163.2330330
[5] S. Bechikh, M. Elarbi, C. Hung, S. Hamdi, and L. B. Said. 2019. A Hybrid Evolutionary Algorithm with Heuristic Mutation for Multi-objective Bi-clustering. In 2019 IEEE Congress on Evolutionary Computation (CEC). IEEE, Wellington, New Zealand, 2323–2330. https://doi.org/10.1109/CEC.2019.8790309
[6] D. Bertsimas and J. Tsitsiklis. 1993. Simulated annealing. Statistical science 8, 1 (1993), 10–15.
[7] J C Bezdek. 2013. Pattern recognition with fuzzy objective function algorithms. Springer Science & Business Media, New York, N. Y.
[8] C. W. Bong and M. Rajeswari. 2011. Multi-objective nature-inspired clustering and classification techniques for image segmentation. Applied Soft Computing 11, 4 (2011), 3271–3282. https://doi.org/10.1016/j.asoc.2011.01.014

[9] C. W. Bong and M. Rajeswari. 2012. Multiobjective clustering with metaheuristic: current trends and methods in image segmentation. IET Image Processing 6, 1 (2012), 1–10. https://doi.org/10.1049/iet-ipr.2010.0122

[10] T. Boongoen and N. Iam-On. 2018. Cluster ensembles: A survey of approaches with recent extensions and applications. Computer Science Review 28 (2018), 1–25. https://doi.org/10.1016/j.cosrev.2018.01.003

[11] M. Boussetelm, S. Bechikh, C. Hung, and L. B. Said. 2017. Bi-MOCK: A Multi-objective Evolutionary Algorithm for Bi-clustering with Automatic Determination of the Number of Bi-clusters. In Neural Information Processing, D. Liu, S. Xie, Y. Li, D. Zhao, and E. M. El-Alfy (Eds.). Springer International Publishing, Cham, 366–376. https://doi.org/10.1007/978-3-319-70899-9_38

[12] M. Brun, C. Sima, J. Hua, J. Lowey, B. Carroll, E. Suh, and E. R. Dougherty. 2007. Model-based Evaluation of Clustering Validation Measures. Pattern Recogn. 40, 3 (March 2007), 807–828. https://doi.org/10.1016/j.patcog.2006.06.026

[13] R. Cheng, Y. Jin, M. Olhofer, and B. Sendhoff. 2016. A Reference Vector Guided Evolutionary Algorithm for Many-Objective Optimization. IEEE Transactions on Evolutionary Computation 20, 5 (2016), 773–791. https://doi.org/10.1109/TEVC.2016.2519378

[14] D. W. Corne, J. D. Knowles, and M. J. Oates. 2000. The Pareto Envelope-Based Selection Algorithm for Multiobjective Optimization. In Parallel Problem Solving from Nature PPSN VI, M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H. Schwefel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 839–848. https://doi.org/10.1007/3-540-45356-3_82

[15] K. Deb, S. Agrawal, A. Pratap, and T. Meyarivan. 2000. A Fast Elitist Non-dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II. In Parallel Problem Solving from Nature PPSN VI, M. Schoenauer, K. Deb, G. Rudolph, X. Yao, E. Lutton, J. J. Merelo, and H. Schwefel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 849–858. https://doi.org/10.1007/3-540-45356-3_83

[16] K. Deb and H. Jain. 2014. An Evolutionary Many-Objective Optimization Algorithm Using Reference-Point-Based Nondominated Sorting Approach. Part I: Solving Problems With Box Constraints. IEEE Transactions on Evolutionary Computation 18, 4 (Aug 2014), 577–601. https://doi.org/10.1109/TEVC.2013.2281535

[17] G N. Demir, A Ş. Uyar, and Ş. Gündüz-Öğüdücü. 2010. Multiobjective evolutionary clustering of web user sessions: a case study in web page analysis. IEEE Transactions on Evolutionary Computation 14, 6 (2010), 579–597. https://doi.org/10.1109/TEVC.2009.2035114

[18] A. G. Di Nuovo, M. Palesi, and V. Catania. 2007. Multi-Objective Evolutionary Fuzzy Clustering for High-Dimensional Problems. In 2007 IEEE International Fuzzy Systems Conference. IEEE, London, UK, 1–6. https://doi.org/10.1109/FUZZY.2007.4295660

[19] Z. Dong, H. Jia, and M. Liu. 2018. An adaptive multiobjective genetic algorithm with fuzzy-means for automatic data clustering. Mathematical Problems in Engineering 2018 (2018), 13 pages. https://doi.org/10.1155/2018/6123874

[20] J. Du, E. E. Korkmaz, R. Alhajj, and K. Barker. 2005. Alternative Clustering by Utilizing Multi-objective Genetic Algorithm with Linked-List Based Chromosome Encoding. In Machine Learning and Data Mining in Pattern Recognition, P.erner and A. Imiya (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 346–355. https://doi.org/10.1007/11510888_34

[21] D. Dutta, P. Dutta, and J. Sil. 2012. Clustering by multi objective genetic algorithm. In 2012 1st International Conference on Recent Advances in Information Technology (RAIT). IEEE, Dhanbad, India, 548–553. https://doi.org/10.1109/RAIT.2012.6194619

[22] D. Dutta, P. Dutta, and J. Sil. 2012. Data clustering with mixed features by multi objective genetic algorithm. In 2012 12th International Conference on Hybrid Intelligent Systems (HIS). IEEE, Pune, India, 336–341. https://doi.org/10.1109/HIS.2012.6421357

[23] D. Dutta, P. Dutta, and J. Sil. 2012. Simultaneous feature selection and clustering for categorical features using multi objective genetic algorithm. In 2012 12th International Conference on Hybrid Intelligent Systems (HIS). IEEE, Pune, India, 191–196. https://doi.org/10.1109/HIS.2012.6421332

[24] D. Dutta, J. Sil, and P. Dutta. 2019. Automatic clustering by multi-objective genetic algorithm with numeric and categorical features. Expert Systems with Applications 137 (2019), 357–379. https://doi.org/10.1016/j.eswa.2019.06.056

[25] P. Dutta and S. Saha. 2017. Fusion of Expression Values and Protein Interaction Information Using Multi-Objective Optimization for Improving Gene Clustering. Comput. Biol. Med. 89, C (Oct. 2017), 31–43. https://doi.org/10.1016/j.compbiomed.2017.07.015

[26] T. Elkaeb and A. Mahmood. 2018. Differential Evolution: A Survey and Analysis. Applied Sciences 8, 10 (2018), 25 pages. https://doi.org/10.3390/app8101945

[27] L. Eröz, M. Steinbach, and V. Kumar. 2002. A New Shared Nearest Neighbor Clustering Algorithm and its Applications. In Workshop on Clustering High Dimensional Data and its Applications at 2nd SIAM International Conference on Data Mining. SIAM, Arlington, VA, USA, 105–115.

[28] K. Faceli, A. C. P. L. F. de Carvalho, and M. C. P. de Souto. 2006. Multi-Objective Clustering Ensemble. In 2006 Sixth International Conference on Hybrid Intelligent Systems (HIS06). IEEE, Rio de Janeiro, Brazil, 51–51. https://doi.org/10.1109/HIS.2006.264934

[29] K. Faceli, M. C. P. de Souto, D. S. A. da Araújo, and A. C. P. L. F. de Carvalho. 2009. Multi-objective Clustering Ensemble for Gene Expression Data Analysis. Neurocomput. 72, 13-15 (Aug. 2009), 2763–2774. https://doi.org/10.1016/j.neucom.2008.09.025

[30] X. Z. Fern and C. E. Brodley. 2004. Solving Cluster Ensemble Problems by Bipartite Graph Partitioning. In Proceedings of the Twenty-First International Conference on Machine Learning. (Banff, Alberta, Canada) (ICML ’04). ACM, New York, NY, USA, 36–https://doi.org/10.1145/1015530.1015414

[31] F. Folino and C. Fizzuti. 2019. A Multiobjective and Evolutionary Clustering Method for Dynamic Networks. In 2010 International Conference on Advances in Social Networks Analysis and Mining. IEEE, Odense, Denmark, 256–263. https://doi.org/10.1109/ASONAM.2010.23

[32] P. Franti, J. Kivi´arvi, T. Kaukoranta, and O. Nevalainen. 1997. Genetic Algorithms for Large-Scale Clustering Problems. Comput. J. 40, 9 (01 1997), 547–554. https://doi.org/10.1093/comjnl/40.9.547 arXiv:https://academic.oup.com/comjnl/article-pdf/40/9/547/981038/400547.pdf

Manuscript submitted to ACM
A Survey of Evolutionary Multi-Objective Clustering Approaches

[33] A. Garcia-Piquer, J. Barcardt, A. Fornells, and E. Golobardes. 2017. Scaling-up multiobjective evolutionary clustering algorithms using stratification. Pattern Recognition Letters 93 (2017), 69–77. https://doi.org/10.1016/j.patrec.2016.12.001

[34] M. Garza-Fabre, J. Handl, and J. Knowles. 2017. A New Reduced-Length Genetic Representation for Evolutionary Multiobjective Clustering. In Evolutionary Multi-Criterion Optimization, H. Trautmann, G. Rudolph, K. Klamroth, O. Schütze, M. Wiecz, Y. Jin, and C. Grimme (Eds.). Springer International Publishing, Cham, 236–251. https://doi.org/10.1007/978-3-319-54157-0_17

[35] M. Garza-Fabre, J. Handl, and J. Knowles. 2018. An Improved and More Scalable Evolutionary Approach to Multiobjective Clustering. IEEE Transactions on Evolutionary Computation 22, 4 (Aug 2018), 515–535. https://doi.org/10.1109/TEVC.2017.2726341

[36] P. Gupta and V. Sharma. 2019. A Survey on Multi-objective based clustering techniques for solving real life problems. In 2019 International Conference on Issues and Challenges in Intelligent Computing Techniques (ICICIT), Vol. 1. IEEE, Ghaziabad, India, 1–4. https://doi.org/10.1109/ICICT46931.2019.8977640

[37] G. Hacioglu, V. F. A. Kand, and E. Sesli. 2016. Multi objective clustering for wireless sensor networks. Expert Systems with Applications 59 (2016), 86–100. https://doi.org/10.1016/j.eswa.2016.04.016

[38] E. Hancor and D. Karaboga. 2017. A comprehensive survey of traditional, merge-split and evolutionary approaches proposed for determination of cluster number. Swarm and Evolutionary Computation 32 (2017), 49–67. https://doi.org/10.1016/j.swevo.2016.06.004

[39] J. Handl and J. Knowles. 2005. Exploiting the Trade-off - The Benefits of Multiple Objectives in Data Clustering. In Evolutionary Multi-Criterion Optimization, C. A. Coello Coello, Arturo Hernández Aguirre, and Eckart Zitzler (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 547–560. https://doi.org/10.1007/114978-3-642-62964-7_4

[40] J. Handl and J. Knowles. 2005. Improvements to the scalability of multiobjective clustering. In 2005 IEEE Congress on Evolutionary Computation, Vol. 3. IEEE, Edinburgh, UK, 2372–2379 Vol. 3. https://doi.org/10.1109/CEC.2005.1554990

[41] J. Handl and J. Knowles. 2005. Multiobjective clustering around medoids. In 2005 IEEE Congress on Evolutionary Computation, Vol. 1. IEEE, Edinburgh, UK, 632–639 Vol. 1. https://doi.org/10.1109/CEC.2005.1554742

[42] J. Handl and J. Knowles. 2007. An Evolutionary Approach to Multiobjective Clustering. IEEE Transactions on Evolutionary Computation 11, 1 (Feb 2007), 56–76. https://doi.org/10.1109/TEVC.2006.877146

[43] J. Handl and J. Knowles. 2012. Clustering Criteria in Multiobjective Data Clustering. In Parallel Problem Solving from Nature - PPSN XII, C. A. C. Coello, V. C. Sel, K. Deb, S. Forrest, G. Nicosia, and M. Pavone (Eds.). Springer, Springer Berlin Heidelberg, Berlin, Heidelberg, 32–41. https://doi.org/10.1007/978-3-642-32964-7_4

[44] J. Handl and J. Knowles. 2013. Evidence Accumulation in Multiobjective Data Clustering. In Evolutionary Multi-Criterion Optimization, Robin C. Purshouse, P. J. Fleming, C. M. Fonseca, S. Greco, and J. Shaw (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 543–557. https://doi.org/10.1007/978-3-642-37140-0_41

[45] J. Horn, N. Nafliotis, and D.E. Goldberg. 1994. A niched Pareto genetic algorithm for multiobjective optimization. In Proceedings of the First IEEE Conference on Evolutionary Computation, IEEE World Congress on Computational Intelligence. IEEE, Orlando, FL, USA, 82–87 vol.1. https://doi.org/10.1109/CEC.1994.350037

[46] E. R. Hruschka and R. J. G. B. Campello and A. Freitas and A. C. P. L. F. de Carvalho. 2009. A Survey of Evolutionary Algorithms for Clustering. IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews) 39, 2 (2009), 133–155. https://doi.org/10.1109/TSMCC.2008.2007252

[47] L. Hubert and P. Arabie. 1985. Comparing partitions. Journal of Classification 2, 1 (01 Dec 1985), 193–218. https://doi.org/10.1007/BF01980875

[48] A. Jain. 2010. Data Clustering: 50 Years Beyond K-Means. Pattern Recognition Letters 31 (06 2010), 651–666. https://doi.org/10.1016/j.patrec.2009.09.011

[49] A. K. Jain and R. C. Dubes. 1988. Algorithms for Clustering Data. Prentice-Hall, Inc., Upper Saddle River, NJ, USA.

[50] A. K. Jain, M. N. Murty, and P. J. Flynn. 1999. Data Clustering: A Review. ACM computing surveys (CSUR) 31, 3 (Sept. 1999), 264–323. https://doi.org/10.1145/331349.331304

[51] M. Kaya and R. Alhaji. 2004. Integrating multi-objective genetic algorithms into clustering for fuzzy association rules mining. In Fourth IEEE International Conference on Data Mining (ICDM04). IEEE, Brighton, UK, 431–434. https://doi.org/10.1109/ICDM.2004.10050

[52] R. A. Khurma and I. Aljarah. 2021. A Review of Multiobjective Evolutionary Algorithms for Data Clustering Problems. Springer Nature, Singapore, 177–199. https://doi.org/10.1007/978-981-33-4191-3_8

[53] O. Kirkland, V. J. Rayward-Smith, and B. de la Iglesia. 2011. A Novel Multi-Objective Genetic Algorithm for Clustering. In Intelligent Data Engineering and Automated Learning - IDEAL 2011, H. Yin, W. Wang, and V. Rayward-Smith (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 317–326. https://doi.org/10.1007/978-3-642-32388-9_38

[54] B. Larsen and C. Aone. 1999. Fast and Effective Text Mining Using Linear-Time Document Clustering. In Proceedings of the Fifth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (San Diego, California, USA) (KDD ’99), Association for Computing Machinery, New York, NY, USA, 16–22. https://doi.org/10.1145/312129.312186

[55] J. S. Lee, S. C. Park, J. J. Lee, and H. H. Ham. 2014. Document clustering using Multi-Objective Genetic Algorithms with parallel programming based on CUDA. In 2014 11th International Conference on Informatics in Control, Automation and Robotics (ICINCO), Vol. 01. IEEE, IEEE, Vienna, Austria, 288–287. https://doi.org/10.1109/ICINCO.2014.6850287

[56] B. Li, J. Li, K. Tang, and X. Yao. 2015. Many-Objective Evolutionary Algorithms: A Survey. Comput. Surveys 48, 1, Article 13 (Sept. 2015), 35 pages. https://doi.org/10.1145/2792884

[57] H. Li, M. Gong, Q. Wang, J. Liu, and L. Su. 2016. A multiobjective fuzzy clustering method for change detection in SAR images. Applied Soft Computing 46 (2016), 767–777. https://doi.org/10.1016/j.asoc.2015.10.044

Manuscript submitted to ACM
[58] J. Li, R. Liu, M. Zhang, and Y. Li. 2017. Ensemble-based multi-objective clustering algorithms for gene expression data sets. In 2017 IEEE Congress on Evolutionary Computation (CEC). IEEE, Donostia, Spain, 333–340. https://doi.org/10.1109/CEC.2017.796931

[59] K. Li, K. Deb, Q. Zhang, and S. Kwong. 2015. An Evolutionary Many-Objective Optimization Algorithm Based on Dominance and Decomposition. IEEE Transactions on Evolutionary Computation 19, 5 (Oct 2015), 694–716. https://doi.org/10.1109/TEVC.2014.2373386

[60] M. Li, S. Yang, and X. Liu. 2014. Shift-Based Density Estimation for Pareto-Based Algorithms in Many-Objective Optimization. IEEE Transactions on Evolutionary Computation 18, 3 (June 2014), 348–365. https://doi.org/10.1109/TEVC.2013.2262178

[61] C. Liu, J. Liu, D. Peng, and C. Wu. 2018. A General Multiobjective Clustering Approach Based on Multiple Distance Measures. IEEE Access 6 (2018), 41706–41719. https://doi.org/10.1109/ACCESS.2018.2860791

[62] C. Liu, Q. Zhao, B. Yan, S. Elsayed, and R. Sarker. 2019. Transfer learning-assisted multi-objective evolutionary clustering framework with decomposition for high-dimensional data. Information Sciences 505 (2019), 449–456. https://doi.org/10.1016/j.patcog.2019.07.099

[63] R. Liu, Y. Liu, and Y. Li. 2012. An improved method for multi-objective clustering ensemble algorithm. In 2012 IEEE Congress on Evolutionary Computation. IEEE, Brisbane, QLD, Australia, 1–8. https://doi.org/10.1109/CEC.2012.6252972

[64] R. Liu, R. Wang, X. Yu, and L. An. 2017. Shape automatic clustering-based multi-objective optimization with decomposition. Machine Vision and Applications 28, 5 (2017), 497–508. https://doi.org/10.1007/s00138-017-0850-6

[65] J. Luo, L. Jiao, and J. Lozano. 2015. A Sparse Spectral Clustering Framework via Multi-Objective Evolutionary Algorithm. IEEE Transactions on Evolutionary Computation 20 (01 2015), 1–1. https://doi.org/10.1109/TEVC.2015.2476359

[66] J. MacQueen. 1967. Some methods for classification and analysis of multivariate observations. In Proceedings of the fifth Berkeley symposium on mathematical statistics and probability, Vol. 1. Oakland, CA, USA, University of California Press, California, USA, 281–297.

[67] K. F. Man, T. M. Chan, K. S. Tang, and S. Kwong. 2004. Jumping-genes in evolutionary computing. In Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (GECCO ’07). Association for Computing Machinery, New York, NY, USA, 861–868. https://doi.org/10.1145/1276958.1277126

[68] N. Matake, T. Hiroyasu, M. Miki, and T. Senda. 2007. Multiobjective Clustering with Automatic K-Determination for Large-Scale Data. In Proceedings of the 9th Annual Conference on Genetic and Evolutionary Computation (London, England) (GECCO ’07). Association for Computing Machinery, New York, NY, USA, 861–868. https://doi.org/10.1145/1276958.1277126

[69] H. D. Menéndez, D. F. Barrero, and D. Camacho. 2013. A multi-objective genetic graph-based clustering algorithm with memory optimization. In 2013 IEEE Congress on Evolutionary Computation. IEEE, IEEE, Cancun, Mexico, 3174–3181

[70] H. D. Menéndez, D. F. Barrero, and D. Camacho. 2014. A Co-Evolutionary Multi-Objective approach for a K-adaptive graph-based clustering algorithm. In 2014 IEEE Congress on Evolutionary Computation (CEC). IEEE, Beijing, China, 2724–2731. https://doi.org/10.1109/CEC.2014.6900369

[71] K. Morik, A. Kaspari, M. Wurst, and M. Skirzynski. 2012. Multi-objective frequent termset clustering. Knowledge and Information Systems 30, 3 (2012), 715–738. https://doi.org/10.1007/s10489-011-0431-3

[72] A. Mukhopadhyay and U. Maulik. 2007. Multiobjective approach to categorical data clustering. In 2007 IEEE Congress on Evolutionary Computation. IEEE, Singapore, 1296–1303. https://doi.org/10.1109/CEC.2007.4424620

[73] A. Mukhopadhyay and U. Maulik. 2009. Unsupervised pixel classification in satellite imagery using multiobjective fuzzy clustering combined with SVM classifier. IEEE transactions on geoscience and remote sensing 47, 4 (2009), 1132–1138. https://doi.org/10.1109/TGRS.2008.2008182

[74] A. Mukhopadhyay, U. Maulik, and C. A. C. Bandyopadhyay, S. andCoello. 2013. Survey of multiobjective evolutionary algorithms for data mining: Part II. IEEE Transactions on Evolutionary Computation 18, 1 (2013), 20–35. https://doi.org/10.1109/TEVC.2013.2290082

[75] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay. 2007. Multiobjective Genetic Fuzzy Clustering of Categorical Attributes. In 20th International Conference on Information Technology (ICT 2007). IEEE, Rourkela, India, 74–79. https://doi.org/10.1109/ICT.2007.13

[76] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay. 2009. Multiobjective Genetic Algorithm-Based Fuzzy Clustering of Categorical Attributes. IEEE Transactions on Evolutionary Computation 13, 5 (Oct 2009), 991–1005. https://doi.org/10.1109/TEVC.2009.2012163

[77] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay. 2010. Simultaneous informative gene selection and clustering through multiobjective optimization. In IEEE Congress on Evolutionary Computation. IEEE, Barcelona, Spain, 1–8. https://doi.org/10.1109/CEC.2010.5886207

[78] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay. 2013. An Interactive Approach to Multiobjective Clustering of Gene Expression Patterns. IEEE Transactions on Biomedical Engineering 60, 1 (2013), 35–41. https://doi.org/10.1109/TBME.2012.2220565

[79] A. Mukhopadhyay, U. Maulik, and S. Bandyopadhyay. 2015. A Survey of Multiobjective Evolutionary Clustering. ACM Computing Surveys (CSUR) 47, 4 (May 2015), 46 pages. https://doi.org/10.1145/2742642

[80] T. Overy and R. Alhaji. 2009. Parallel clustering of high dimensional data by integrating multi-objective genetic algorithm with divide and conquer. Applied Intelligence 31, 3 (2009), 318. https://doi.org/10.1007/s10489-008-0129-8

[81] M. K. Pakhira, S. Bandyopadhyay, and U. Maulik. 2004. validity index for crisp and fuzzy clusters. Pattern Recognition 37, 3 (2004), 487–501. https://doi.org/10.1016/j.patcog.2003.06.005

[82] A. K. Paul and P. C. Shill. 2018. New automatic fuzzy relational clustering algorithms using multi-objective NSGA-II. Information Sciences 448-449 (2018), 112–133. https://doi.org/10.1016/j.ins.2018.03.025

[83] A. Pirzada, H. R. Mashhadi, and S. Hamed Javadi. 2013. An optimal energy-efficient clustering method in wireless sensor networks using multi-objective genetic algorithm. International Journal of Communication Systems 26, 1 (2013), 114–126. https://doi.org/10.1002/dac.1336

[84] J. Piantoni, K. Facelli, T. C. Sakata, J. C. Pereira, and M. C. P. de Souto. 2015. Impact of base partitions on multi-objective and traditional ensemble clustering algorithms. In International Conference on Neural Information Processing. Springer, Springer, Cham, Switzerland, 696–704. https://doi.org/10.1007/978-3-319-26532-2_77

Manuscript submitted to ACM
A Survey of Evolutionary Multi-Objective Clustering Approaches

[85] C. Pizzuti and A. Socievole. 2019. Multiobjective Optimization and Local Merge for Clustering Attributed Graphs. *IEEE transactions on cybernetics* 50, 12 (2019), 4997–5009. https://doi.org/10.1109/TCYB.2018.2889413

[86] K. Praditwong, M. Harman, and X. Yao. 2010. Software module clustering as a multi-objective search problem. *IEEE Transactions on Software Engineering* 37, 2 (2010), 264–282. https://doi.org/10.1109/TSE.2010.26

[87] X. Qian, X. Zhang, L. Jiao, and W. Ma. 2008. Unsupervised texture image segmentation using multiobjective evolutionary clustering ensemble algorithm. In *2008 IEEE Congress on Evolutionary Computation (IEEE World Congress on Computational Intelligence)*. IEEE, Hong Kong, China, 3561–3567. https://doi.org/10.1109/CEC.2008.4632179

[88] M. A. Rahman and M. Z. Islam. 2014. A hybrid clustering technique combining a novel genetic algorithm with K-Means. *Knowledge-Based Systems* 71 (2014), 345–365. https://doi.org/10.1016/j.knosys.2014.08.011

[89] S. Rana, S. Jasola, and P. Kumar. 2011. A review on particle swarm optimization algorithms and their applications to data clustering. *Artificial Intelligence Review* 35, 3 (2011), 211–222.

[90] W. M. Rand. 1971. Objective Criteria for the Evaluation of Clustering Methods. *J. Amer. Statist. Assoc.* 66, 336 (1971), 846–850. https://doi.org/10.1080/01621459.1971.10482356

[91] K. S. N. Ripon and M. N. H. Siddique. 2009. Evolutionary multi-objective clustering for overlapping clusters detection. In *2009 IEEE Congress on Evolutionary Computation*, IEEE, Trondheim, Norway, 976–982. https://doi.org/10.1109/ICEC.2009.4983051

[92] K. S. N. Ripon, C. Tsang, S. Kwong, and Man-Ki I. 2006. Multi-Objective Evolutionary Clustering using Variable-Length Real Jumping Genes Genetic Algorithm. In *18th International Conference on Pattern Recognition (ICPR*6), Vol. 1. IEEE, Hong Kong, China, 1200–1203. https://doi.org/10.1109/ICPR.2006.827

[93] K. S. N. Ripon, Chi-Ho Tsang, and S. Kwong. 2006. Multi-Objective Data Clustering using Variable-Length Real Jumping Genes Genetic Algorithm and Local Search Method. In *The 2006 IEEE International Joint Conference on Neural Network Proceedings*, IEEE, Vancouver, BC, Canada, 3609–3616. https://doi.org/10.1109/IJCNN.2006.416214

[94] N. Riquelme, C. Von Lucken, and B. Baran. 2015. Performance metrics in multi-objective optimization. In *2015 Latin American Computing Conference (CLEI)*. IEEE, Arequipa, Peru, 1–11. https://doi.org/10.1109/CLEI.2015.7360024

[95] R. C. Romero-Zaliz, C. Rubio-Escudero, J. P. Cobb, F. Herrera, O. Cordón, and I. Zaliz. 2008. A multiobjective evolutionary conceptual clustering methodology for gene annotation within structural databases: a case of study on the gene ontology database. *IEEE Transactions on Evolutionary Computation* 12, 6 (2008), 679–701. https://doi.org/10.1109/TEVC.2008.915995

[96] S. Saha and S. Bandyopadhyay. 2013. A generalized automatic clustering algorithm in a multiobjective framework. *Applied Soft Computing*, 13, 1 (2013), 89–108. https://doi.org/10.1016/j.asoc.2012.06.005

[97] H. Schütze, C. D. Manning, and P. Raghavan. 2008. *Introduction to information retrieval*. Vol. 39. Cambridge University Press Cambridge, Cambridge, England.

[98] O. C. Sert, K. Dursun, T. Özyer, J. Jida, and R. Alhajj. 2012. The Unification and Assessment of Multi-Objective Clustering Results of Categorical Datasets with H-Confidence Metric. *J. UCS* 18, 12 (2012), 507–531.

[99] O. C. Sert, K. Dursun, and T. Özyer. 2011. Ensemble of Multi-objective Clustering Unified with H-Confidence Metric as Validity Metric. In *2011 International Conference on Advances in Social Networks Analysis and Mining*, IEEE, Kaohsiung, Taiwan, 537–541. https://doi.org/10.1109/ASONAM.2011.95

[100] R. Shang, H. Liu, and L. Jiao. 2017. Multi-objective clustering technique based on k-nodes update policy and similarity matrix for mining communities in social networks. *Physica A: Statistical Mechanics and its Applications* 466 (2017), 1–24. https://doi.org/10.1016/j.physa.2017.05.026

[101] J. Shi and J. Malik. 2000. Normalized cuts and image segmentation. *IEEE Transactions on pattern analysis and machine intelligence* 22, 8 (2000), 888–905. https://doi.org/10.1109/34.886888

[102] S. Shirakawa and T. Nagao. 2009. Evolutionary image segmentation based on multiobjective clustering. In *2009 IEEE Congress on Evolutionary Computation*, IEEE, Trondheim, Norway, 2466–2473. https://doi.org/10.1109/ICEC.2009.4983250

[103] P. Siarry. 2016. *Metaheuristics*. Springer, Cham, Switzerland. https://doi.org/10.1007/978-3-319-45403-0

[104] A. Skabar and K. Abdalgader. 2013. Clustering Sentence-Level Text Using a Novel Fuzzy Relational Clustering Algorithm. *IEEE Transactions on Knowledge and Data Engineering* 25, 1 (2013), 62–75. https://doi.org/10.1109/TKDE.2011.205

[105] P. H. A. Sneath. 1957. The application of computers to taxonomy. *Microbiology* 17, 1 (1957), 201–226. https://doi.org/10.1099/00221287-17-1-201

[106] R. R Sokal. 1958. A statistical method for evaluating systematic relationships. *Univ. Kansas, Sci. Bull.* 38 (1958), 1409–1438.

[107] J. Sørensen. 1948. *A Method of Establishing Groups of Equal Amplitude in Plant Sociology Based on Similarity of Species Content and Its Application to Analyses of the Vegetation on Danish Commons*. I kommission hos E. Munksgaard, København, Denmark. https://books.google.com.br/books?id=rpS8GAAACAAJ

[108] R. Srikanth, R. George, N. Waris, D. Prabhu, F. E. Petry, and B. P. Buckles. 1995. A Variable-Length Genetic Algorithm for Clustering and Classification. *Pattern Recogn. Lett.* 16, 8 (Aug. 1995), 789–800. https://doi.org/10.1016/0167-8655(95)00043-G

[109] A. Steudel. 2002. *Relationship-based Clustering and Cluster Ensembles for High-dimensional Data Mining*. Ph.D. Dissertation. The University of Texas. doi: 10.1109/IJCNN.2006.247372

[110] S. Tan, M. Steinbach, and V. Kumar. 2013. Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining* (2013), 487–533.

[111] J. Timmis, A. Hone, T. Stibor, and E. Clark. 2008. Theoretical advances in artificial immune systems. *Theoretical Computer Science* 403, 1 (2008), 11–32. https://doi.org/10.1016/j.tcs.2008.02.011

Manuscript submitted to ACM
[111] C. Tsai, W. Chen, and M. Chiung. 2012. A modified multiobjective EA-based clustering algorithm with automatic determination of the number of clusters. In 2012 IEEE International Conference on Systems, Man, and Cybernetics (SMC). IEEE, Seoul, Korea (South), 2833–2838. https://doi.org/10.1109/ICSMC.2012.6378178

[112] A. Wahid, X. Gao, and P. Andreae. 2015. Multi-objective clustering ensemble for high-dimensional data based on Strength Pareto Evolutionary Algorithm (SPEA-II). In 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, Paris, France, 1–9. https://doi.org/10.1109/DSAA.2015.7344795

[113] L. Wang, G. Cui, Q. Zhou, and K. Li. 2020. A multi-clustering method based on evolutionary multiobjective optimization with grid decomposition. Swarm and Evolutionary Computation 55 (2020), 100691. https://doi.org/10.1016/j.swevo.2020.100691

[114] Q. Wang, H. Li, M. Gong, L. Su, and L. Jiao. 2014. A multiobjective optimization method based on MOEA/D and fuzzy clustering for change detection in SAR images. In 2014 IEEE Congress on Evolutionary Computation (CEC). IEEE, Beijing, China, 3024–3029. https://doi.org/10.1109/CEC.2014.6902659

[115] Y. Wang, J. A. Walker, S. J. Bale, M. A. Trefzer, and A. M. Tyrrell. 2015. Two-phase multiobjective genetic algorithm for constrained circuit clustering on FPGAs. In 2015 IEEE Congress on Evolutionary Computation (CEC). IEEE, Sendai, Japan, 1183–1190. https://doi.org/10.1109/CEC.2015.7257023

[116] S. Zhu, L. Xu, and L. Cao. 2018. A Study of Automatic Clustering Based on Evolutionary Many-Objective Optimization. In Proceedings of the Genetic and Evolutionary Computation Conference Companion (Kyoto, Japan) (GECCO ’18). Association for Computing Machinery, New York, NY, USA, 69–76. https://doi.org/10.1145/3205651.3205759

[117] H. Xia, J. Zhuang, and D. Yu. 2013. Novel Soft Subspace Clustering with Multi-Objective Evolutionary Approach for High-Dimensional Data. Pattern Recogn. 46, 9 (Sept. 2013), 2562–2575. https://doi.org/10.1016/j.patcog.2013.02.005

[118] Y. Zhao, C. Li, H. Liu, R. Lan, and C. W. Chen. 2018. Noise robust multiobjective evolutionary clustering image segmentation motivated by the intuitionistic fuzzy information. IEEE Transactions on Fuzzy Systems 27, 2 (2018), 387–401. https://doi.org/10.1109/TFUZZ.2018.2852289

[119] E.-C. Wei and C.-K. Cheng. 1991. Ratio cut partitioning for hierarchical designs. IEEE Transactions on Computer-Aided Design of Integrated Circuits and Systems 10, 7 (1991), 911–921. https://doi.org/10.1109/43.67601

[120] F. Zhao, J. Fan, H. Liu, R. Lan, and C. W. Chen. 2018. Noise robust multiobjective evolutionary clustering image segmentation motivated by the intuitionistic fuzzy information. IEEE Transactions on Fuzzy Systems 27, 2 (2018), 387–401. https://doi.org/10.1109/TFUZZ.2018.2852289

[121] H. Zhao, J. Liang, and C. Deng. 2017. Clustering Ensemble Selection for Categorical Data Based on Internal Validity Indices. Pattern Recogn. 69, C (Sept. 2017), 150–168. https://doi.org/10.1016/j.patcog.2017.04.019

[122] Y. Zheng, L. Jia, and H. Cao. 2012. Multi-objective gene expression programming for clustering. Information Technology and Control 41, 3 (2012), 283–294. https://doi.org/10.1007/s10786-012-0289-9

[123] D. Zhi, Y. Yao, and L. Wang. 2020. A multi-objective genetic algorithm with fuzzy c-means for automatic data clustering. Applied Soft Computing 24 (2014), 679 – 691. https://doi.org/10.1016/j.asoc.2014.08.036

[124] J. Zhu, G. Cui, Q. Zhou, and K. Li. 2020. A multi-clustering method based on evolutionary multiobjective optimization with grid decomposition. Swarm and Evolutionary Computation 55 (2020), 100691. https://doi.org/10.1016/j.swevo.2020.100691

[125] P. Andreae. 2015. Multi-objective clustering ensemble for high-dimensional data based on Strength Pareto Evolutionary Algorithm (SPEA-II). In 2015 IEEE International Conference on Data Science and Advanced Analytics (DSAA). IEEE, Paris, France, 1–9. https://doi.org/10.1109/DSAA.2015.7344795

[126] Y. Zhu, L. Xu, and L. Cao. 2018. Many-objective fuzzy centroids clustering algorithm for categorical data. Expert Systems with Applications 103 (2018), 283–294. https://doi.org/10.1016/j.eswa.2018.02.013

[127] K. Li, G. Cui, Q. Zhou, and L. Wang. 2020. A multi-clustering method based on evolutionary multiobjective optimization with grid decomposition. Swarm and Evolutionary Computation 55 (2020), 100691. https://doi.org/10.1016/j.swevo.2020.100691

Manuscript submitted to ACM