Iraqi Journal for Computer Science and Mathematics

Journal Homepage: http://journal.esj.edu.iq/index.php/IJCM
e-ISSN: 2788-7421 p-ISSN: 2958-0544

Convex Optimization Techniques for High-Dimensional Data Clustering Analysis: A Review

Ahmed Yacoub Yousif [1,2,]*, Basad Al-Sarray [3]*

[1] University of Baghdad, College of Science, Mathematics Department, Baghdad, Iraq
[2] University of Technology, Department of Applied Sciences, Baghdad, Iraq
[3] University of Baghdad, College of Science, Computer Science Department, Iraq

*Corresponding Author: Ahmed Yacoub Yousif, Basad Al-Sarray

DOI: https://doi.org/10.52866/ijcsm.2024.05.03.022
Received February 2024; Accepted May 2024; Available online July 2024

ABSTRACT: Clustering techniques have been instrumental in discerning patterns and relationships within datasets in data analytics and unsupervised machine learning. Traditional clustering algorithms struggle to handle real-world data analysis problems where the number of clusters is not readily identifiable. Moreover, they face challenges in determining the optimal number of clusters for high-dimensional datasets. Consequently, there is a demand for enhanced, adaptable and efficient techniques. Convex clustering, rooted in a rich mathematical framework, has steadily emerged as a pivotal alternative to traditional techniques. It amalgamates the strengths of conventional approaches while ensuring robustness and guaranteeing globally optimal solutions. This review offers an in-depth exploration of convex clustering, detailing its formulation, challenges and practical applications. It examines synthetic datasets, which serve as foundational platforms for academic exploration, emphasizing their interactions with the semi-smooth Newton augmented Lagrangian (SSNAL) algorithm. Convex clustering provides a robust theoretical foundation, but challenges, including computational limitations with expansive datasets and noise management in high-dimensional contexts, persist. Hence, the paper discusses current challenges and prospective future directions in the domain. This research aims to illuminate the potency and potential of convex clustering in modern data analytics, highlighting its robustness, flexibility and adaptability across diverse datasets and applications.

Keywords: Convex clustering, Unsupervised learning, High-dimensional data, Regularization, Global optimality, Semi-smooth Newton, Augmented Lagrangian algorithm.

INTRODUCTION

Clustering is an essential tool in data science research and is widely used in various applications, from image processing [1] to artificial intelligence [2] and genomics [3]. Particularly in genomics, it identifies groups of genes with similar patterns of expression, which may indicate functional relationships, particularly in genomics [4]. One of the solutions to the sparse of data is optimization algorithms, which are widely used and implemented in high-dimensional space. These algorithms have been Improvement of optimization algorithms directed towards nature behaviour to efficiently mimic optimization solutions [5], [6].

Clustering, as illustrated in Figure 1, groups a set of data objects into clusters to maximize the similarity between objects in the same cluster and the dissimilarity between objects in the different clusters [7]. Several clustering techniques have been proposed and implemented, with partition-based techniques being the most prevalent [8]. These methods, including the ubiquitous k-means [9], are characterized by their reliance on cluster prototypes and a predetermined number of clusters, denoted as 'k'.

Several clustering techniques have been proposed and implemented like k-means have been instrumental in numerous applications, they suffer from inherent challenges, including sensitivity to initialization, vulnerability to local minima [10], [11] and the requirement of specifying k a priori. Several strategies have been employed to circumnavigate these challenges, such as innovative initialization techniques [12], [13], theoretical model alterations [14] and heuristic-based adjustments [15], [16]. However, these approaches have limitations, notably when confronted with non-uniform, nonconvex or skewed datasets. Additionally, with the surge in volume and complexity of data, traditional clustering methodologies have grappled with their inherent limitations [17]. The curse of dimensionality [18], a phenomenon where data becomes sparse in high-dimensional spaces, rendering the distances between points less meaningful, poses significant challenges [19].

*Corresponding author: author@organization.edu.co

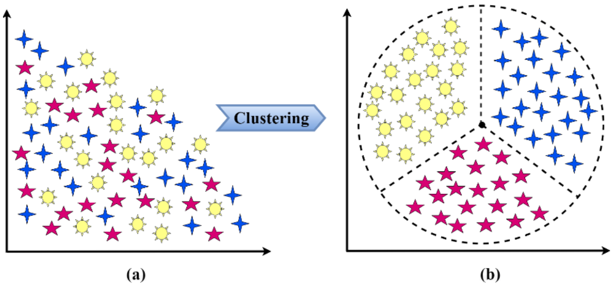


FIGURE 1. - A partition with n = 65 and k = 3.

The convex clustering model has emerged as a solution to tackle the problems related to prototype selection and optimization [20]. This approach is known for its nature and provides a theoretical assurance of finding global minima. Moreover, there has been progress, in the development of the semi-smooth Newton augmented Lagrangian (SSNAL) method [21] aimed at handling large scale issues in convex clustering.

Our exploration, there is still a lack of understanding regarding the structures of local minima. This paper aims to bridge this gap by providing an, in depth analysis of synthetic datasets and conducting an exploration of the SSNAL algorithm and also emphasize the significance of convex clustering in data analytics.

This investigation delves on an exploration of the following:

1. A detailed examination of synthetic datasets and their inherent structures.
2. The influence of gamma values on cluster dynamics.
3. A comparative analysis of convex clustering compared to traditional methods underscores its robustness, flexibility and unique emphasis on the data structure.

The subsequent sections of this paper are structured as follows: Section 2 explores our research methodology and offers insights into our dataset statistics; Section 3 presents an overview of prevalent clustering algorithms; Section 4 delves into the formulation of the convex clustering problem; Section 5 examines various convex optimization techniques tailored for clustering; Section 6 elucidates the general steps for deploying convex clustering techniques; Section 7 presents a detailed examination of synthetic datasets; Section 8 reflects on the challenges in the realm of convex clustering and outlines future research directions; and Section 9 concludes this paper with a discussion.

1. RESEARCH METHODOLOGY AND DATA STATISTICS

This research's methodology and data statistics are grounded in bibliometric data, encompassing detailed information about a publication, such as its authors, keywords and document type. The common indices for this bibliometric data are Scopus, Web of Science (WoS) and Google Scholar. We have utilized multiple databases for this study to ensure a comprehensive selection of high-quality publications spanning various disciplines.

Our research focuses on the nuances of convex clustering, specifically emphasizing its intricate relationship with optimization and algorithmic techniques. Consequently, our keyword search palette ranged from convex clustering and clustering algorithms to data mining and convex optimization.

Our search spanned from 1996 to 2023 to encompass the significant developmental phase of convex clustering, starting from its foundational stages. This period marks the emergence of key research and seminal works in the field, providing a complete historical-to-contemporary overview. The distribution of these documents by year is illustrated in Figure 2.

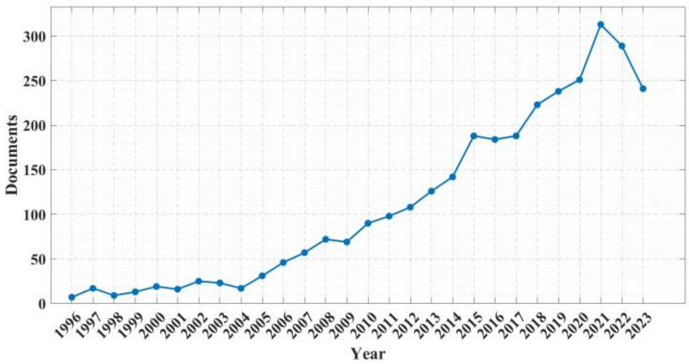


FIGURE 2. - Distribution of documents by year.

In addition to analysing the temporal distribution, we also examined the geographical spread of research contributions. The bar chart in Figure 3 visually compares the document counts from various countries, highlighting the global interest and diverse contributions in the field of convex clustering. This geographical analysis underscores the widespread academic engagement with convex clustering worldwide, with countries like China and the United States leading in document counts.

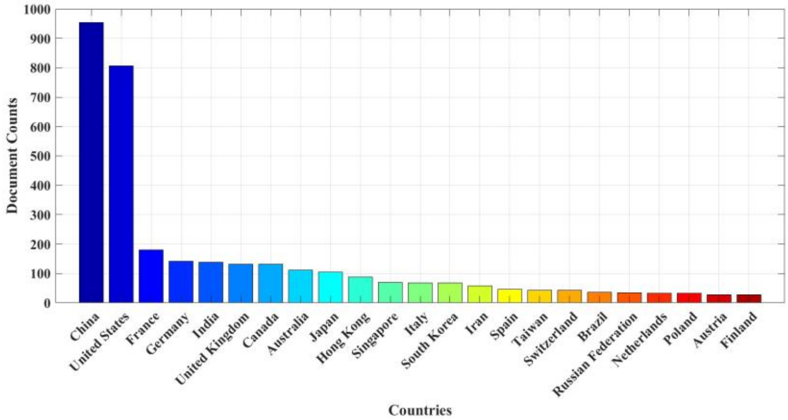


FIGURE 3. - Document counts for various countries.

Further, an analysis by subject area reveals that computer science, mathematics and engineering are the predominant fields contributing to the research on convex clustering. Figure 4 illustrates the proportional contributions of these and other disciplines, showcasing the multi-disciplinary nature of research in this area. The diversity of disciplines reflects the wide range of applications and interest in convex clustering within the academic community.

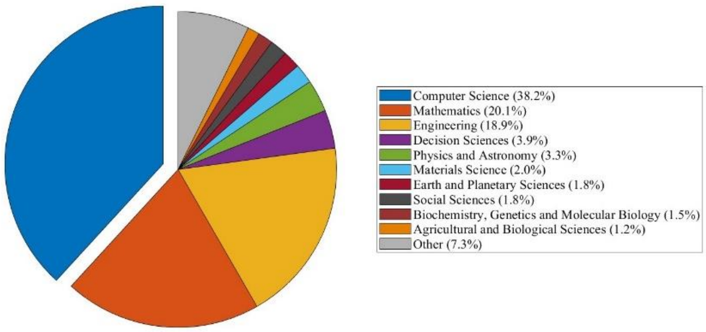


FIGURE 4. - Proportional contributions of different disciplines.

2. OVERVIEW OF CLUSTERING ALGORITHMS

Clustering is a versatile method used to identify and group similar data. Its significance has surged due to the increasing volume of data, leading to a pressing need for refined data segmentation strategies. The k-means algorithm [22] is notable for its simplicity and efficacy compared to various other algorithms. The essence of clustering revolves around assessing similarities, often quantified using mathematical distance measures. Jain and Dubes (1988) articulate this concept by highlighting that members within a cluster should closely resemble each other while markedly different from those outside their group. It is essential that the criteria for these assessments are both clear and actionable [23], [24]. Broadly, clustering algorithms can be categorized into several groups (as shown in Figure 5).

- **Partition-Based Clustering [25]:** Partition-based clustering divides data into k clusters, each represented by a central point, which can be either a centroid (the average location of all points) or a medoid (the most representative point). The prominent algorithms in this category are k-means and k-medoid, renowned for their efficiency in clustering data based on distance.
- **Hierarchical Clustering (HC) [26]:** Hierarchical clustering (HC) categorizes data objects into hierarchical levels, forming iterative clusters using either top-down or bottom-up approaches, allowing for data exploration at different granularity levels. The agglomerative method merges single objects into larger clusters, while the divisive method breaks down clusters until each object forms a single cluster or a stopping criterion is met.
- **Density-Based Clustering [27], [28]:** Density-based clustering identifies dense regions as clusters, with high-density modes forming a cluster centre and sparse areas marked as noise and outliers. This method enables the clustering of arbitrary shapes and effectively eliminates outliers or noisy data points, as seen in algorithms like a density-based clustering algorithm (DBSCAN) and ordering points to identify the clustering structure (OPTICS).
- **Model-Based Clustering [29], [30]:** This clustering approach assumes that data originates from a specific probability distribution or model, with each component corresponding to a distinct cluster. The most common approach to model-based clustering is the Gaussian mixture model (GMM). The primary goal is to enhance the

model's alignment with the observed data. A notable advantage is its ability to pinpoint outliers. However, it necessitates the estimation of model parameters, often employing the Expectation-Maximization (EM) algorithm as a central method.

- **Spectral and Graph-Based Clustering [31]:** Spectral clustering (SC) is a type of clustering algorithm that forms groups based on the connections between data points. It forecasts the data into a lower-dimensional space using the eigenvalues and eigenvectors of the data matrix. The concept of a data graph representation underpins it, with nodes representing data points and edges denoting similarities between data points.
- **Biclustering or Co-Clustering [32]–[34]:** Biclustering is a versatile analytical tool that clusters rows and columns of a matrix to identify checkerboard-like biclusters. It is flexible enough to identify subgroup structures in both dimensions, offering unique advantages in gene expression analysis. Systematically clustering genes and samples can accurately identify local structures, such as in plaid mode (PM) and spectral co-clustering.

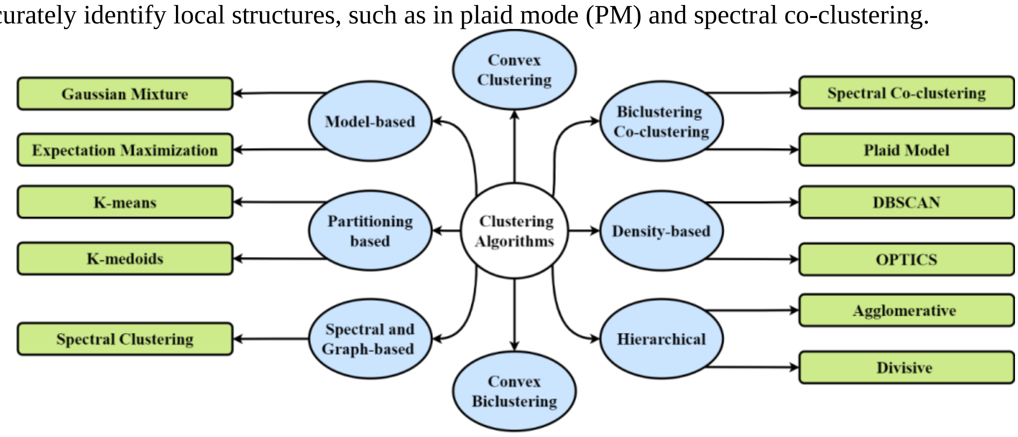


FIGURE 5. - A graphical representation illustrating the diverse categories of clustering algorithms.

High-dimensional spaces pose challenges that traditional single-objective clustering techniques may struggle with. Different data sets may require varied validity measures, and the efficiency of algorithms decreases as cluster overlap increases [35], [36]. Table 1 below provides a concise summary of various clustering algorithms.

Advancements, in research have greatly contributed to our knowledge and practical implementation of clustering algorithms in the areas of evaluating cluster validity and determining optimal clusters. Li et al. [37] proposed a method for evaluating cluster validity by using the ratio of the deviation of sum-of-squares and Euclidean distance. They demonstrated its effectiveness in dynamically identifying near-optimal cluster numbers across diverse datasets. Chowdhury et al. [38] developed an entropy-based initialization method for the k-means algorithm, enhancing cluster determination in multidimensional image datasets. Fang et al. [39] contributed to this field by introducing the clustering deviation index (CDI), a novel metric for evaluating clustering accuracy in single-cell RNA sequencing (scRNA-seq) data. Challenging the conventional unsupervised approach of clustering algorithms, Sinaga and Yang [40] presented a procedure for k-means that autonomously calculates the optimal number of clusters, broadening its applicability to various datasets. Safari et al. [41] innovatively refined the bisecting k-means algorithm to automate the selection of cluster quantity, thereby achieving notable enhancements in computational efficiency and resource utilization. Yin et al. [42] adeptly employed mixed-data clustering techniques in the context of life insurance risk management. They focused on critically analysing discrepancies between expected and actual death claims by applying the k-prototype clustering method. Al-Janabee and Al-Sarray [43] explored the clustering of brain tumour gene expression data using fuzzy c-means (FCM) and its hybrid forms combined with particle swarm optimization (PSO) and a genetic algorithm (GA). In the context of internet of things (IoT) and big data, Li et al. [44] proposed the meta-clustering ensemble scheme based on model selection (MCEMS), leveraging agglomerative hierarchical clustering (AHC). Uykan [45] extended the EM formulation for GMM by creating a fusion of centroid-based and graph clustering termed hybrid-nongreedy asynchronous clustering (H-NAC). This approach outperformed k-means, SC and structured graph learning (SGL). These developments underscore the growing sophistication and practical relevance of clustering techniques in data analysis.

Convex clustering has emerged as a notable solution, merging the advantages of traditional techniques with robustness, the ability to guarantee global optimal solutions [46], [47] and flexibility without needing a predefined number of clusters [48]. Drawing upon a foundation [49], [50] it is important to acknowledge that this approach does have its limitations. These limitations encompass issues, with the quality of data representation, computational hurdles when dealing with datasets [51], the need for re-optimization with dynamic data and challenges in handling noise in high-dimensional scenarios [52].

Understanding of clustering techniques, especially when working with data that has high-dimensions, is extremely important. The emergence and significance of convex clustering methods highlight this point. The subsequent sections will further explore these techniques, examining their theoretical foundations and practical applications.

Table 1. - An overview of various clustering algorithms.

Algorithm	Mathematical Formula	No. Eq.	Noise Handling	Cluster Requirements	Scalability for High Dimension	*Ref.*
k-means	$min_S \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2$	(1)	Poor	Yes	Poor	[53], [54]
HC	$min_{x \in X, y \in Y}\, dist\,(x, y)$	(2)	Poor	No	Poor	[8], [55]
DBSCAN	$min_S (dist\,(x_i, S) \le \epsilon)$	(3)	Good	No	Poor	[56], [57]
EM	$\sum_{k=1}^{K} \pi_k f_k (A; \theta_k)$	(4)	Moderate	No	Poor	[29], [58]
SC	$min_Y\, tr\,(Y^T L Y)$	(5)	Poor	Yes	Poor	[31], [59]
PM	$X = \sum_{l=1}^{L} \left(a_l \cdot b_l^T\right) + E$	(6)	Moderate	Yes	Moderate	[60], [61]

3. CONVEX CLUSTERING PROBLEM FORMULATION

Convex clustering has become increasingly popular as an efficient alternative to conventional clustering techniques such as k-means. However, because the K-Means algorithm is NP-hard and k has to be given first, convex clustering offers a more practical solution [62], [63]. Lindsten et al. [64] proposed convex relaxed k-means clustering, which uses a fusion penalty to achieve agglomerative clustering. Also, [65] offers an exhaustive analysis of convex clustering, delineating various model formulations and optimization strategies. This comprehensive review expands both the theoretical and practical understanding of convex clustering, highlighting its statistical properties and its multifaceted applications across diverse fields.

Basically, to control convex clustering via data alongside with the cluster's shapes. This is achievable by adjusting hyperparameter denoted λ. This hyperparameter is responsible for the cluster centroids positions and it potentially affect the performance efficacy of the cluster. Choosing the right λ ensure avoiding trivial results solutions. Several research studies, including those by Tan and Witten [66], Chi et al. [67] and Wang et al. [52], have delved into the nuances of selecting λ. The concept is further developed in weighted convex clustering, where weights are judiciously assigned to the regularization term. This adjustment enhances stability against outliers and noise [68]. Despite its mathematical allure, convex clustering requires pristine data features, making it susceptible to challenges in noisy scenarios. In this context, it is noteworthy that Prony's method, a widely used method for modelling signals using a finite sum of exponential terms, has been explored for potential improvements through nuclear-norm-penalized regularization to enhance its stability against noise perturbations [69]. Such complementary techniques could offer valuable insights for addressing the vulnerability of convex clustering to noise in high-dimensional datasets.

The mathematical model of convex clustering, expressed of the sum of norms, can be formulated in various ways, see Table 2 below. Consider a dataset represented by data points $x_1, x_2, \dots, x_n$, where n represents the number of observations and d indicates the number of features for each observation. Each observation x_i stands for a distinct entity, and the features capture various attributes or characteristics associated with the entity.

The standard convex clustering approach is fundamentally structured as follows:

$$\min_{C,A} \sum_{i=1}^{n} \|x_i - c_{a_i}\|_2^2 + \lambda \sum_{i,j} \|c_i - c_j\|_{p_n} \tag{7}$$

Where p_n can take values from the set $\{1,2,+\infty\}$. The hyperparameter γ plays a pivotal role in dictating the clustering's granularity, as higher values consolidate data into fewer clusters.

The weighted version of convex clustering, designed to offset the potential penalization of centroids, is articulated as follows:

$$\min_{C,A} \sum_{i=1}^{n} \|x_i - c_{a_i}\|_2^2 + \lambda \sum_{i,j} w_{ij} \|c_i - c_j\|_{p_n} \tag{8}$$

The weights, w_{ij}, are computed based on the proximity of neighbouring data points x_i and x_j, using the function defined as:

$$w_{p,q} = \begin{cases} \exp\left(-\phi \|x_i - x_j\|^2\right) & \text{if } i,j \in \epsilon \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

Where: $\epsilon = \bigcup_{i=1}^{n} \{(i,j) \mid j$ is one of the k-nearest neighbours of i and $i < j \le n\}$ and $\phi > 0$ is a constant.

Hocking et al. [68] emphasized that the distance between data points primarily influences the selection of weights. In many of the experiments presented in the paper, the weights decay based on the distance, following the formula $w_{ij} = \exp\left(-\|X_i - X_j\|^2/2\right)$. This approach suggests that points closer to each other are assigned higher weights than those that are farther apart. Furthermore, in certain scenarios, a uniform weight of $w_{ij} = 1$ is adopted to streamline computation. The central objective behind this weight selection strategy is to guarantee accurate cluster formation and avert undesired cluster splits. This effectively shapes the solution path's geometry and determines the final number of clusters tailored to the dataset.

Table 2. - Overview of various convex clustering formulations.

Item	Description	No. Eq.	Ref.
Formulation	$\min_{X\in\mathbb{S}^n} \operatorname{Tr}(AX)$ s.t, $X \geq 0,\ X_{ii} = 1\forall i,\ \operatorname{rank}(X) = k$	(10)	[70]
Details	where A is the matrix of squared Euclidean distances between the data points, X is the Gram matrix of the data points, $\operatorname{Tr}(AX)$ is the trace of AX, and k is the number of clusters.		
Remarks	Semi-definite programming (SDP) relaxation of k-means, aiming to minimize the trace of AX. The efficiency of this approach depends on the available SDP solvers.		
Formulation	$\min_{C\in\mathbb{R}^{n\times p}} \frac{1}{2} \parallel X - C \parallel_F^2 + \gamma \parallel DBC \parallel_{1,q}$	(11)	[71]
Details	D is a diagonal matrix with weights w_{ij}. B is a sparse matrix. C represents the cluster centers matrix.		
Remarks	Tree-structured convex clustering (TCC) uses a tree structure to represent data as a graph. It promotes cluster formation by summing the l_q norms of the DBC matrix's rows.		
Formulation	$\min_{C_i} \sum_{i=1}^{m} \left(\operatorname{trace}(C_i^T L_i) + \beta\lVert C_i\rVert_1\right) + \frac{\alpha}{2}\sum_{\substack{1\leq i,j\leq m\\ i\neq j}} \lVert C_i - C_j\rVert_F^2$	(12)	[72]
Details	The term trace $(C_i^T L_i)$ reduces within-cluster variation using the similarity matrix C_i and Laplacian $L_i \cdot \beta\lVert C_i\rVert_1$ promotes C_i sparsity, guided by positive $\beta.\frac{\alpha}{2}\sum_{1\leq i,j\leq m, i\neq j} \parallel C_i - C_j \parallel_F^2$ that ensure similarity across views, adjusted by positive α.		
Remarks	Convex relaxation of the pairwise sparse spectral clustering (PSSC) model can be solved more efficiently and used to compute the solution matrices for the nonconvex model.		
Formulation	$\min_C \sum_i \lVert S_i(C_i - X_i)\rVert^2 + \lambda \sum_i \sum_j w_{ij}\lVert C_i - C_j\rVert^2$	(13)	[73]
Details	C_i is the centre for the i th data point X_i. S_i weights the features of X_i and, for missing data, indicates observed features. λ is the regularization factor, and w_{ij} weights the distance influence between points i and j on cost.		
Remarks	Convex relaxation of the pairwise sparse spectral clustering (PSSC) model can be solved more efficiently and used to compute the solution matrices for the nonconvex model.		
Formulation	$\min_{C,E} \parallel X - P - E \parallel_F^2 + \alpha \sum_{i<j} w_{ij}\lVert c_i - c_j\rVert_p + \lambda \parallel E \parallel_{2,1}$	(14)	[74]
Details	The term $\parallel X - P - E \parallel_F^2$ represents the fidelity term for the difference between data points and the combined clustering and robust components. The regularization term remains unaltered. The expression $\lambda \parallel E \parallel_{2,1}$ regularizes E for row sparsity, which is useful for detecting outlier features. In this context, λ indicates the data noise level.		
Remarks	Robust convex clustering (RCC) extends the base model to handle outliers and missing data. It decomposes data into clustering and robust components using a $l_{2,1}$-norm to enforce row-wise sparsity and identify outliers.		
Formulation	$\min_{C,B} \frac{1}{2}\sum_{j=1}^{N} \left(x_j - c_j\right)^T B\left(x_j - c_j\right) + \gamma \sum_{1\leq j_1<j_2\leq N} w_{j_1 j_2}\lVert c_{j_1} - c_{j_2}\rVert_1$	(15)	[75]
Details	x_j and c_j are the j th data point and its cluster center. B is a positive definite matrix for the learned metric. $w_{j_1 j_2}$ weighs the connection between points j_1 and $j_2 \cdot \gamma$ balances fidelity and regularization. The condition $\log\det(B) > 0$ ensures B is positive definite, maintaining convexity and metric relevance.		
Remarks	The convex clustering with metric learning (CCML) method extends the standard algorithm by incorporating a positive definite matrix B to weigh the features based on their relevance and noise levels.		
Formulation	$\min_C \frac{1}{2} \parallel X - C \parallel_2^2 + \gamma \sum_{i<j} w_{ij}\lVert c_i - c_j\rVert_q + \lambda \sum_{j=1}^{d} w_j \lVert c_j\rVert_2$	(16)	[52]
Details	$\parallel X - C \parallel_F^2$ is the fidelity term reflecting error between data and clusters. $\sum_{i<j} w_{ij} \parallel c_i - c_j \parallel_2$ regularizes with w_{ij} weights indicating similarity and measuring cluster center distances. $\sum_{j=1}^{d} w_j \lVert c_j\rVert_2$ is the group-lasso penalty, using w_j weights, promoting sparsity in C' 's columns, indicating less informative		

Item	Description	No. Eq.	Ref.
	features. γ and λ adjust the balance among the terms.		
Remarks	Sparse convex clustering method, which extends standard convex clustering by incorporating a group-lasso penalty to encourage sparsity in the columns of the solution matrix C. This sparsity helps exclude non-informative features from the clustering.		
Formulation	$\min_{C} \frac{1}{2} \parallel X - C \parallel_F^2 + \gamma \sum_{i<j} w_{ij} \lVert c_i - c_j \rVert_2 + \lambda \sum_{j=1}^{d} \left((1-\rho) w_j \lVert c_j \rVert_2 + \rho \lVert c_j \rVert_1 \right)$	(17)	[76]
Details	C is the matrix of cluster centres, where each row is a centre. Vector w signifies weights, while γ, λ, are tuning parameters for balancing fidelity, group lasso and lasso penalties. The fidelity term, $\parallel X - C \parallel_F^2$, measures the data-cluster discrepancy. $\sum_{i<j} w_{ij} \lVert c_i - c_j \rVert_2$ is the regularization of w_{ij} indicating similarity and measuring cluster distance. The sparse group lasso penalty, $\sum_{j=1}^{d} \left((1-\rho) w_j \lVert c_j \rVert_2 + \rho \lVert c_j \rVert_1 \right)$, encourages group and individual sparsity, adjusted by ρ.		
Remarks	The sparse group lasso convex clustering (SGLCC) method encourages fidelity to the data and the formation of clusters. The second term is a sparse group lasso penalty that encourages sparsity in the columns of C (variable selection) and within-group sparsity.		
Formulation	$\min_{C} \frac{1}{2} \sum_{i=1}^{N} w_i \lVert x_i - c_i \rVert_2^2 + \gamma \sum_{i<j} w_{ij} \lVert c_i - c_j \rVert_2$	(18)	[77]
Details	Each data point has a nonnegative weight w_i, while pairs have weights w_{ij}. In this setup, x_i is the i th data point, and c_i its cluster center. γ is a nonnegative parameter balancing fidelity and regularization.		
Remarks	Sum-of-norms (SON) clustering model with multiplicative weights. It is an extension of the SON clustering model that incorporates multiplicative weights $w_i$into the objective function.		
Formulation	$\min_{C} \frac{1}{2} \parallel X - C \parallel_F^2 + \lambda \left(\sum_{i<j} w_{ij} \lVert c_i - c_j \rVert_2 + \sum_{k<l} \tilde{w}_{kl} \lVert c_k - c_l \rVert_2 \right)$	(19)	[78]
Details	The matrix X holds the data, and C represents clusters. $\parallel X - C \parallel_F^2$ measures error between data and clusters. $\sum_{i<j} w_{ij} \lVert c_i - c_j \rVert_2$ and $\sum_{k<l} \tilde{w}_{kl} \lVert c_k - c_l \rVert_2$ are regularization terms with w_{ij} and $\tilde{w}_{kl}$ as similarity weights for rows and columns. λ balances fidelity and regularization.		
Remarks	The convex biclustering model formulation simultaneously clusters patients and genetic pathways by fusing the columns and rows in the data matrix.		
Formulation	$\min_{C \in \mathbb{R}^{n_1 \times n_2 \times n_3}} \frac{1}{2} \parallel X - C \parallel_2^2 + \gamma R(C)$	(20)	[67]
Details	The Frobenius norm, $\lVert \cdot \rVert_F$, measures the difference between tensors X and C, pushing C to closely match $X. \gamma$ is a nonnegative tuner, with larger values emphasizing a checker box pattern. $R(C)$ is regularization term promoting a checker box pattern by comparing tensor slices. These terms urge similar slices to merge, with the merge strength determined by weights $w_{d,ij}$. Higher weights push slices closer.		
Remarks	The convex objective function $F_\gamma(C)$ plays a pivotal role in the co-clustering approach by driving the estimation process. The objective of this approach is to reduce the value of this function in order to approximate the provided tensor while simultaneously achieving the intended checker box clustering pattern.		

4. CONVEX OPTIMIZATION TECHNIQUES FOR CLUSTERING

In the realm of data analytics and machine learning clustering techniques have attracted interest as they provide sophisticated tools, for understanding and unravelling complex datasets. Out of the techniques available convex clustering and biclustering have become particularly popular because of their inherent reliability and ability to generate globally optimal solutions.

The inception of the lasso, introduced by Tibshirani in 1996 [79], marked a pivotal moment as it instigated sparse coefficients within linear models, striking an optimal balance between subset selection and ridge regression. This pioneering method laid the groundwork for Pelckmans et al. in 2005 [80], who explored convex optimization for clustering. They ingeniously incorporated a shrinkage term, which led to sparse centroids. Tibshirani et al. [81] further enhanced this research in 2005, introducing the concept of a fused lasso to emphasize both coefficient sparsity and local constancy, especially in high-dimensional datasets.

In 2011, Boyd and his colleagues emphasized the potential of the alternating direction method of multipliers (ADMM), for solving distributed optimization problems. This approach is particularly important when dealing with large scale challenges [82]. Around the time Lindsten and his team introduced a method called sum of norms (SON) which is an adaptive approach, for determining the number of clusters. On they proposed a modified version of k-means clustering that utilizes optimization techniques [20] [64].

Over the years, researchers have ardently explored and enhanced convex clustering. A testament to this is the work of Zhu et al. in 2014 [83], who conducted an exhaustive theoretical analysis of the SON clustering technique. Their insights validated the efficacy of it under certain specified conditions. Another noteworthy contribution was made by Chi and Lange in 2015 [84], who conceptualized two innovative splitting methods for convex clustering. Their comparative analysis of the efficacies of ADMM and the alternating minimization algorithm (AMA) revealed the latter's superiority. In the same year, Tan and Witten [66] embarked on exploring the statistical intricacies of convex clustering. Their findings unveiled its interrelationships with other clustering techniques and elucidated the optimal range for its tuning parameter. Recent work has assessed nuclear-norm-penalized regularization's potential to enhance Prony's method's stability in signal modelling, especially with noisy, low-rank signals [69].

The intricate web of research in convex clustering and biclustering, encompassing foundational works and recent advancements, is visually represented in Figure 6. Each node epitomizes a unique study, with the node size denoting its citation frequency. The interconnections, indicated by arrows, showcase the evolution of ideas and how contemporary research builds upon prior foundations. The chronological timeline at the base provides a temporal context, emphasizing the burgeoning growth and escalating interest in this domain over time.

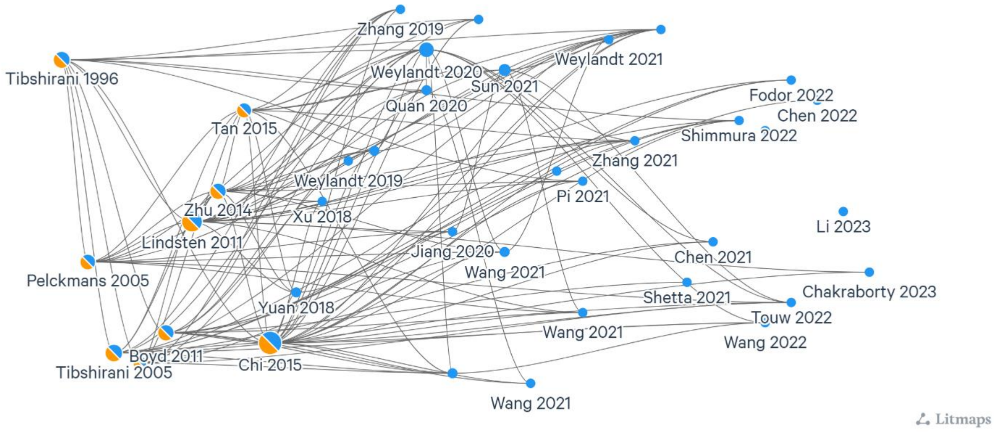


FIGURE 6. - Citation network of convex clustering and biclustering research.

Convex clustering's allure is undeniable. However, its efficacy can be contingent upon the dataset's size and intricacy. To mitigate these challenges, scholars have proposed an array of strategies, including the SSNAL method, adaptive sieving (AS), enhanced AS (EAS) techniques, the majorization penalty algorithm and other ADMM-centric methods. The viability of these approaches has been empirically tested across diverse datasets, spanning from simulated and synthetic data to classic ones like MNIST, Iris and WINE, and even specialized datasets like The cancer genome atlas (TCGA) datasets.

Table 3 provides a detailed summary of recent research endeavours in this domain from 2018 to the present. This table delineates the specific challenges addressed by each study, the datasets they employ, their strategies and their pivotal findings. By scrutinizing these studies, we aim to provide a panoramic view of the current landscape and the prospective trajectories of convex clustering and biclustering.

Table 3. - Convex clustering and biclustering: An analysis of methodologies, challenges and datasets used.

Author, Year	Term	Analysis
Yuan et al. (2018) [85]	Problem	Solve large-scale problems with existing algorithms for convex clustering.
	Dataset(s)	Synthetic, MNIST, Fisher Iris, WINE and Yale face B datasets.
	Remark and Key Findings	Semismooth Newton based augmented Lagrangian method for large-scale convex clustering problems. SSNAL shows superior performance and scalability compared to the state-of-the-art software CVXCLUSTR.
Xu et al. (2018) [86]	Problem	Limited applicability and restrictiveness of existing total variation based convex clustering.
	Dataset(s)	Synthetic, Iris, AR face, LIBRAS movement and leaf datasets.
	Remark and	A weighted sum-of-ℓ_1-norm relating convex model. The proposed convex model

Author, Year	Term	Analysis
	Key Findings	has better empirical performance compared to standard clustering methods.
Weylandt, M. (2019) [78]	Problem	The computational properties of convex formulations of biclustering and tensor co-clustering have been less well studied.
	Dataset(s)	Presidential speeches and TCGA breast cancer datasets.
	Remark and Key Findings	Three efficient operator-splitting methods for the convex co-clustering problem are presented: a standard two-block ADMM, a generalized ADMM is far more efficient for large problems.
Zhang, M. (2019) [87]	Problem	The instability issue is a significant drawback of traditional nonconvex clustering methods, often leading to inconsistent clustering results.
	Dataset(s)	Synthetic and Iris datasets.
	Remark and Key Findings	Forward-stagewise clustering for convex clustering. Forward-stagewise clustering can correctly identify the underlying clusters.
Zhou et al. (2020) [88]	Problem	Convex clustering approach with current state-of-the-art algorithms requiring large computation and memory space.
	Dataset(s)	Synthetic Gaussian, Mouse embryo scRNA-seq and Human PBMC scRNA-seq datasets.
	Remark and Key Findings	The smoothing proximal gradient algorithm (Sproga) outperforms ADMM- or AMA-based convex clustering algorithms in speed by one to two orders of magnitude. It requires at least an order of magnitude less memory. Additionally, it surpasses popular algorithms like k-means and hierarchical clustering.
Weylandt et al. (2020) [51]	Problem	Despite the promise of convex clustering, it has not been widely adopted due to its computational intensity and lack of compelling visualizations.
	Dataset(s)	Genomics and text analysis datasets.
	Remark and Key Findings	The convex clustering via algorithmic regularization paths (CARP), which employs algorithmic regularization, offers a 100-fold speedup over conventional methods, provides a more refined approximation grid and enhances clustering solution visualization.
Quan, Z., & Chen, S. (2020) [47]	Problem	Convex clustering method is sensitive to outliers, causing skewed clustering results.
	Dataset(s)	Synthetic, USPS digit and UCI datasets.
	Remark and Key Findings	A robust convex clustering algorithm (RCC) retains the benefits of CC, such as convexity, global optimality and initialization stability, while excelling in outlier detection and clustering quality.
Jiang, T., & Vavasis, S. (2020) [77]	Problem	Validating cluster assignments from various algorithm approximations and boosting the efficacy of sum-of-norms clustering.
	Dataset(s)	A mixture of Gaussians and simulated half moons dataset.
	Remark and Key Findings	The effectiveness of sum-of-norms clustering was shown to be enhanced with multiplicative weights. Several areas for further investigation were identified, including the test's complexity and the generation of weights without prior data information.
Chen et al. (2020) [76]	Problem	The instability of classical clustering methods is due to their tendency to sink into the local optimal solutions of the nonconvex optimization model.
	Dataset(s)	Simulated, brain A, SRBCT, prostate, colon and ARB datasets.
	Remark and Key Findings	The sparse group lasso convex clustering (SGLCC), which autonomously determines cluster counts and group data while filtering nonessential features, was introduced. Using the semi proximal alternating direction method of multipliers (sPADMM) for implementation, it exhibited enhanced performance and feature selection capabilities.
Chen, J., & Suzuki, J. (2021) [89]	Problem	The standard convex clustering problem contains a non-differentiable function optimization, which is inefficient.
	Dataset(s)	Presidential speeches, TCGA and DLBCL datasets.
	Remark and Key Findings	The proposed method combines ALM with Nesterov's accelerated gradient method, improving efficiency and stability across a range of λ values. It also opens avenues for further improvements and application to other clustering problems.
Zhang et al. (2021) [90]	Problem	Convex clustering is noted for its performance and global optimality guarantees, but it's high computational cost makes it challenging for large datasets.
	Dataset(s)	Simulated, lympho, gene, Frey faces, RNA, anuran, fashion and Kuzushiji-MNIST datasets.

Author, Year	Term	Analysis
	Remark and Key Findings	Dynamic programming was used to tackle L1 convex clustering. The introduced C-PAINT method visualizes cluster paths. With uniform weights in L1 convex clustering, this efficient algorithm surpassed others and addressed computational challenges, enabling full cluster path recovery for extensive datasets.
Weylandt et al. (2021) [91]	Problem	Simultaneous denoising and clustering of noisy signals.
	Dataset(s)	Synthetic signal and NMR spectroscopy datasets.
	Remark and Key Findings	A sparse convex wavelet clustering technique merges wavelet denoising with sparse convex clustering. It surpasses current algorithms, offering global solutions and clearer, wavelet-sparse cluster centroids, enhancing result interpretation and data compression.
Wang et al. (2021)[92]	Problem	Enhancing the efficiency of the sum-of-norms (SON) model for clustering.
	Dataset(s)	Simulated, Iris, WINE, letter-recognition, knowledge and MNIST datasets.
	Remark and Key Findings	Euclidean distance matrix model based on the SON model, majorization penalty algorithm. The proposed model and majorization penalty algorithm demonstrate high efficiency in solving the SON model for clustering.
Wang et al. (2021) [93]	Problem	High-dimensional sparse clustering with compositional data.
	Dataset(s)	Simulated and central nervous system (CNS) gene expression datasets.
	Remark and Key Findings	Compositional convex clustering with sparse group lasso (CCC-SGL), employs a proximal gradient descent within the ADMM framework, enhancing clustering by filtering excess features and picking cluster-specific traits.
Wang, M., & Allen, G. I. (2021) [94]	Problem	Integrative clustering of high-dimensional mixed multi-view data.
	Dataset(s)	Simulated, text mining and genomics datasets.
	Remark and Key Findings	The integrative generalized convex clustering optimization (iGecco) method uses an adaptive shifted group lasso penalty and a generalized multi-block ADMM algorithm. It picks optimal features from each data view, enhancing integrative clustering.
Sun et al. 2021 [21]	Problem	Theoretical guarantee and an efficient algorithm for the general weighted convex clustering model.
	Dataset(s)	Synthetic, MNIST, Fisher Iris, WINE and Yale face B datasets.
	Remark and Key Findings	Semismooth Newton-based augmented Lagrangian method. The proposed algorithm has superior performance and scalability, solving a convex clustering problem with 200,000 points in R3 in about six minutes.
Shimamura, K., & Kawano, S. (2021) [95]	Problem	Sparse convex clustering depends heavily on the data and can reduce estimation accuracy when the sample size is insufficient.
	Dataset(s)	Generated two half-moons and LIBRAS movement datasets.
	Remark and Key Findings	Based on Bayesian lasso and global-local shrinkage priors, the proposed Bayesian sparse convex clustering method, with Gibbs sampling, delivers accurate MAP estimation and addresses weight dependencies in the regularization term.
Pi et al. (2021) [96]	Problem	Clustering tasks often pose the challenge of nonconvex problems without clear global optima. The objective was to craft a rapid-converging algorithm for convex clustering issues.
	Dataset(s)	Simulated, synthetic, Iris, seeds, mammal and modified lung cancer gene datasets.
	Remark and Key Findings	A dual formulation for convex clustering uses a first-order project gradient method, line search, adaptive restart and a regularization framework. This method quickly converges, accurately clusters by leveraging sparsity in the fusion penalty, and performs better with a sparsity-inducing norm in high dimensions.
Shetta et al. (2022) [97]	Problem	Challenges in multi-view clustering methods for multi-omic data. Existing methods are computationally complex and do not consider the intrinsic manifold structure of the data.
	Dataset(s)	Simulated datasets and TCGA breast, esophageal, endometrioid, kidney renal clear cell and lung squamous cell carcinoma datasets.
	Remark and Key Findings	Convex graph regularized multi-view clustering method that is robust to outliers. The method is compared to state-of-the-art convex and nonconvex multi-view and single view clustering methods, particularly for clustering cancer subtypes. The results show superior performance in integrating different views by considering the complementary information present in each view. The method

Author, Year	Term	Analysis
		also demonstrated a better ability to discover cancer subtypes than other state-of-the-art multi-view methods.
Chen et al. (2022) [98]	Problem	Clustering algorithms often struggle with heavy-tailed data. The new robustification parameter adds complexity to optimal parameter selection.
	Dataset(s)	Simulated and lung cancer gene expression datasets.
	Remark and Key Findings	A robust convex biclustering (RCBC) version integrates Huber loss and an automatic tuning-free method for optimal robustification parameter selection. It greatly surpasses COBRA in heavy tail situations and streamlines the parameter optimization process.
Yuan et al. (2022) [99]	Problem	Accelerate algorithms for large-scale convex optimization with inherent structured sparsity.
	Dataset(s)	Simulated and MNIST datasets.
	Remark and Key Findings	The adaptive sieving (AS) and enhanced AS (EAS) techniques boost the SSNAL algorithm's speed by over seven times and the ADMM algorithm by over 14 times.
Touw et al. (2022) [100]	Problem	Convex clustering's limitations with large data sets and hierarchical structure issues.
	Dataset(s)	Simulated, banknote, musk, MAGIC telescope and MNIST datasets.
	Remark and Key Findings	A new, efficient algorithm (CCMM) is proposed to minimize the convex clustering loss function and compute the cluster path. Outperforms AMA and SSNAL in speed tests.
Shimmura, R., & Suzuki, J. (2022) [101]	Problem	Efficiency challenges in sparse estimation problems such as fused lasso and convex clustering.
	Dataset(s)	Data generated by a Gaussian distribution.
	Remark and Key Findings	The method transforms ADMM solutions into proximal gradient ones, enhancing efficiency for sparse estimation. It is especially promising for cases with two additional regularization terms.
Fodor et al. (2022) [102]	Problem	Scalable parallel and distributed solvers for convex clustering are limited.
	Dataset(s)	Synthetic data sets and Iris data.
	Remark and Key Findings	The introduced parallel distributed ADMM-based convex clustering algorithm scales well in clusters, efficiently handles large datasets and may surpass existing methods in speed and accuracy.
Armacki et al. (2022) [103]	Problem	Personalized federated learning lacks automatic user model clustering without prior knowledge of cluster structure.
	Dataset(s)	Simulated binary data.
	Remark and Key Findings	The proposed algorithm generalizes convex clustering for automatic model clustering, personalization and generalization, efficiently solved using PDMM. Numerical experiments confirm effectiveness.
Wang et al. (2023) [104]	Problem	Existing biclustering algorithms require extra smoothing steps to obtain informative biclusters and cannot incorporate the compositional constraints required for certain data.
	Dataset(s)	Simulated murine gut microbiome and murine microbiome datasets.
	Remark and Key Findings	A new algorithm for the standard convex biclustering and its extension under the compositional constraints (bi-ADMM and biC-ADMM) can provide a clear checkerboard-like pattern directly without any further smoothing step. biC-ADMM can solve biclustering problems with compositional constraints.
Chakraborty, S., & Xu, J. (2023) [105]	Problem	High-dimensional data challenges convex clustering because of vague pairwise affinities and less effective Euclidean fit metrics.
	Dataset(s)	Simulated low SNR, LIBRAS movement and leukaemia datasets.
	Remark and Key Findings	Biconvex clustering optimizes feature weights with centroids, tackling dimensionality issues and reducing reliance on tuned heuristics. It enhances feature selection during clustering and improves quality.
Li et al. (2023) [106]	Problem	The k-means algorithm easily gets stuck in spurious local minima, and the number of clusters has to be given a priori.
	Dataset(s)	Synthetic, HTRU2, Iris, WINE, X8D5K and Statlog datasets.
	Remark and Key Findings	Multi-prototypes convex merging based k-means clustering algorithm (MCKM), integrating MPS and CM, eliminates the need to pre-specify cluster count and achieves a superior local minimum for the k-means issue.

Additionally, through Table 3, we aim to offer valuable insights into the evolving landscape of convex clustering and biclustering, the challenges researchers encounter, and the innovative solutions being developed to tackle these challenges.

5. GENERAL STEPS FOR CONVEX CLUSTERING TECHNIQUES

Convex clustering, as previously elucidated, amalgamates the principles of k-means clustering with regularization. This fusion optimally balances the categorization of data points while preserving desired attributes, such as smooth cluster assignments. The following steps provide a structured blueprint for effectively implementing convex clustering techniques (see Figure 7):

1. **Data Preprocessing:** Prior to clustering, it is imperative to normalize or standardize the dataset, ensuring feature uniformity. Given the inherent challenges of high-dimensionality, techniques such as principal component analysis (PCA) [107] or t-distributed stochastic neighbor embedding (t-SNE) [108] can expedite the clustering process and potentially elevate its quality.
2. **Problem Mathematical Formulation:** The core objective function for convex clustering aims to minimize the within-cluster distances [84], ensuring that data points within a cluster are as close to each other as possible. Additionally, incorporating a convex penalty further enhances the robustness of the solution.
3. **Convex Optimization Methods:** The efficacy of convex clustering hinges on advanced optimization techniques. Central to this is the ADMM [82], which adeptly navigates the challenges of convexity and variable splitting inherent in the shrinkage term [84]. A key development is the method for estimating cluster matrices in Gaussian mixture using semi-definite programming. This technique employs a Bregman-ADMM-type algorithm centred on pairwise distances, integrating embedding and clustering via pairwise affinity analysis [109]. The AMA transforms unconstrained problems into their constrained, nonsmooth variants, broadening its scope through sparse iteration [83]. First-order strategies include stochastic splitting [63] and the Frank-Wolfe methods [68]. Second-order methods like SSNAL [21] also harmonize efficiency and precision through the nuanced handling of smooth, non-linear equations.
4. **Evaluation:** Three established metrics are employed to assess the performance of clustering algorithms: the F-measure (F^*), normalized mutual information (NMI) and adjusted Rand index (ARI) [110]–[112]. These metrics quantify the concordance between the clustering outcomes and the ground truth.
5. **Interpretation:** The culmination of the clustering process is the profound interpretation of the derived clusters. Valuable insights can be gleaned by examining data attributes within clusters or analysing the cluster centroids' characteristics [113].

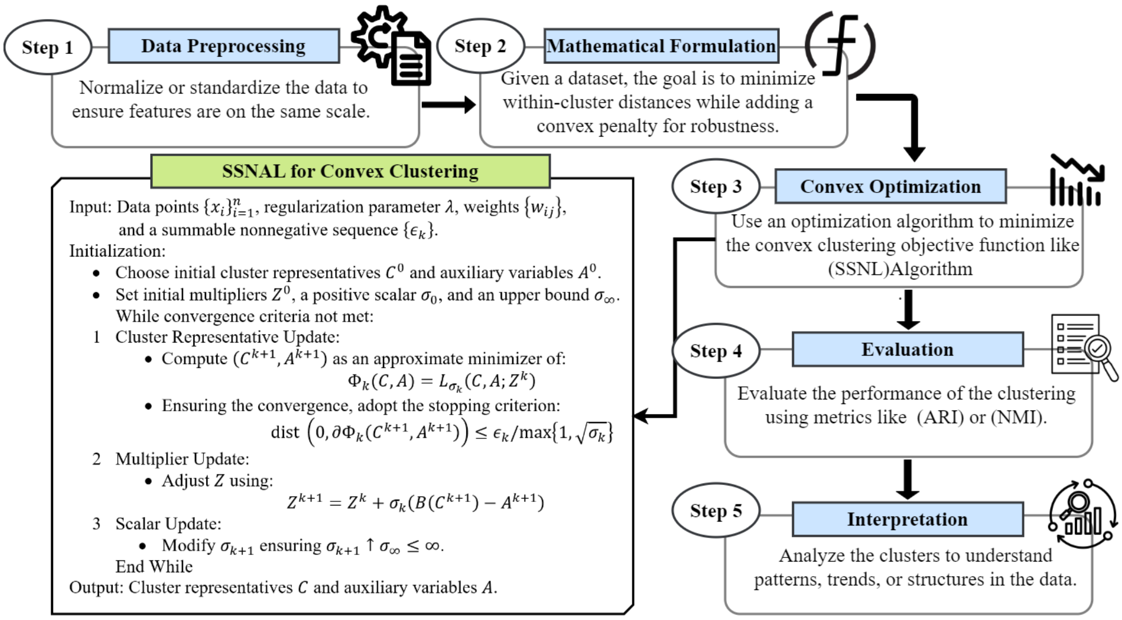


FIGURE 7. - Structured procedure for employing convex clustering techniques.

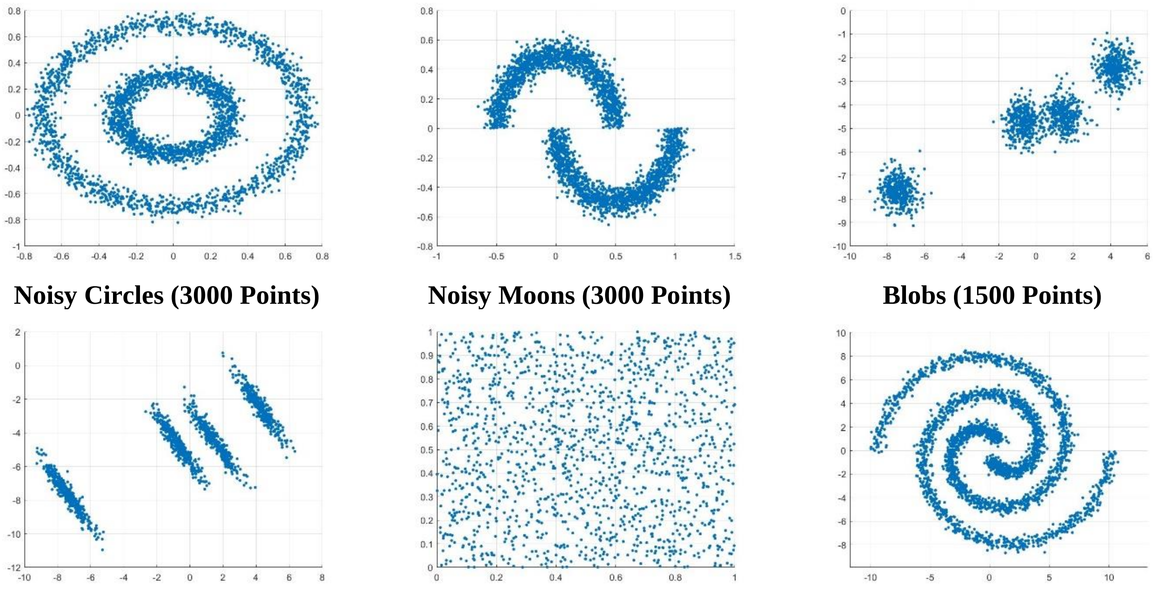
6. EXAMINATION OF SYNTHETIC DATASETS

Synthetic datasets have become indispensable in data science and analytics, serving as essential tools to rigorously test algorithms, establish benchmarks and foster academic exploration. Specifically tailored to emulate real-world scenarios, these datasets provide a controlled environment for algorithm validation and refinement. In this review, we thoroughly examined six selected synthetic datasets. We explored their intrinsic structures, scrutinized their interactions with the SSNAL [21] algorithm and underscored the overarching probability of convex clustering in contemporary data analytics.

6.1 VISUALIZATION OF SYNTHETIC DATASETS: EXPLORING INTRINSIC STRUSTURES

Understanding the inherent geometries and patterns of our chosen datasets is paramount. Detailed scatter plots offer a vivid snapshot of each dataset's landscape. As showcased in Figure 8, these datasets include variations such as unbalanced data, nonconvex configurations and convex shapes. The synthetic datasets encompass:

- **Noisy Circles:** The dataset called noisy circles consists of two circles that are positioned inside each other with some random points scattered between them. This dataset creates difficulties for clustering algorithms that depend heavily on measuring distances because the inner and outer circles do not have a separation, between their points.
- **Noisy Moons:** The dataset has a resemblance to two crescent moons. Consists of overlapping semi circles with some added noise. Its unique structure poses difficulties for approaches that assume clusters, with convex shapes.
- **Blobs:** This dataset consists of a few clusters or blobs that are approximately spherical, in shape. Although this arrangement may appear simple the presence of noise and different densities can make the clustering process more complicated.
- **Anisotropic Blobs:** This dataset is a version of the typical blobs stretching them into elongated clusters. It serves as a test for algorithms to determine their ability to handle clusters with shapes rather, than just the usual spherical ones. Basic algorithms designed for clusters may face challenges when dealing with this dataset.
- **No Structure:** The arrangement of points on this canvas seems random without any clear pattern or purpose. It poses a challenge, for algorithms to avoid fitting closely or attempting to categorize non-existent groups.
- **Twin Spiral:** This complex dataset includes a pair of spirals that wrap around a centre. The intertwined pattern of the spirals poses a challenge, for clustering algorithms to differentiate between the two separate structures.



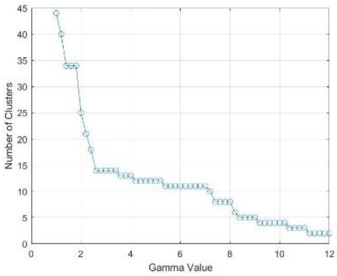
Noisy Circles (3000 Points) **Noisy Moons (3000 Points)** **Blobs (1500 Points)**

Anisotropic (1500 Points) **No Structure (1500 Points)** **Twin Spiral (3000 Points)**

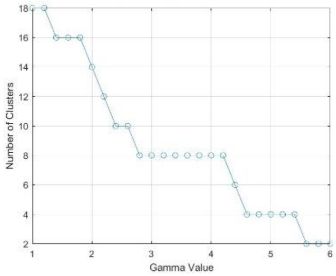
FIGURE 8. - Synthesized data structures and their visual representation.

6.2 GAMMA VALUES AND THEIR BEARING ON CLUSTER DYNAMICS

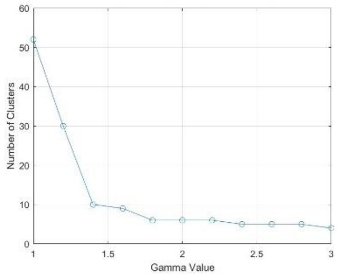
The gamma parameter plays a role in the SSNAL algorithm. It works like a dial that adjusts the sensitivity and responsiveness of the algorithm to data structures. By tweaking gamma, we can achieve outcomes in terms of clustering from highly segmented groups to more broadly categorized clusters. To gain an understanding of how the algorithm behaves we carefully plotted the number of clusters identified at different gamma values. Figure 9 provides a representation that highlights the algorithms stability and adaptability across diverse datasets. The graph demonstrates the versatility of the SSNAL algorithm showcasing its ability to strike a balance, between segmentation and accurate cluster identification.



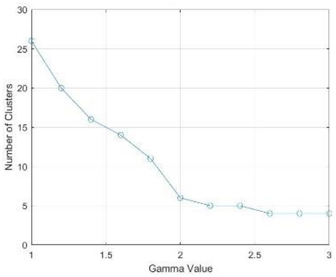
Noisy Circles
($\gamma \in [1, 12]$, $k_n = 10$)



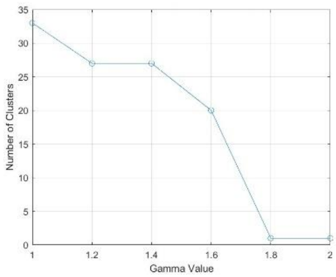
Noisy Moons
($\gamma \in [1,35]$, k_n =15)



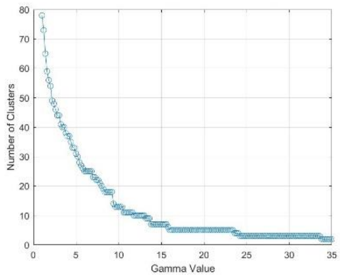
Blobs
($\gamma \in [1,3]$, $k_n = 10$)



Anisotropic
($\gamma \in [1, 3]$, $k_n = 15$)



No Structure
($\gamma \in [1,2]$, $k_n = 10$)



Twin Spiral
($\gamma \in [1,35]$, $k_n = 30$)

FIGURE 9. - Interplay of γ values and cluster dynamics across data types.

6.3 THE PARADIGM OF CONVEX CLUSTERING: A COMPARATIVE ANALYSIS

Convex clustering has been receiving interest because of its ability to effectively detect clusters and encourage both cohesion within clusters and separation, between clusters. This study provides a comparative analysis of convex clustering, specifically the SSNAL algorithm, in comparison to established methods such as k-means clustering, SC, GMM, DBSCAN and convex clustering. Further details and visual representations can be found in Figure 10.

When analysing the **Noisy Circles** dataset, k-means faced challenges due to its inherent assumption of spherical clusters. In contrast, both SC and GMM effectively identified the circular structures. The DBSCAN, a density-based method, misclassified some data points as noise.

K-means once again struggled to effectively cluster the **Noisy Moons** dataset, while SC aptly captured the moon shapes. GMM was slightly less effective, experiencing issues at the moon intersections, whereas DBSCAN was mostly successful.

K-means, SC and GMM were all effective in clustering the **Blobs** dataset, which presents roughly spherical clusters. However, DBSCAN was more conservative, marking some data as noise. On the **Anisotropic Blobs** dataset, K-means had difficulty with non-spherical clusters. SC and GMM excelled, but DBSCAN flagged many points as noise.

Regarding the **No Structure** dataset, both k-means and GMM tried to impose a structure that did not exist. DBSCAN recognized the lack of a clear density-based structure and classified most data points as noise.

The **Twin Spiral** dataset posed a unique challenge. k-means and GMM failed to capture the intertwined spirals, while SC was successful. DBSCAN was mostly effective but identified some regions as noise.

Convex clustering offers an alternative approach that bridges k-means and HC. It clusters data while penalizing differences in cluster centroids, allowing flexibility in shaping clusters. Convex clustering showed potential in handling noisy circles and noisy moons datasets, but its performance depended on the regularization parameter's tuning. The dataset of blobs exhibited performance to that of k-means. On the hand both SC and DBSCAN showcased their capabilities when it came to the twin spiral dataset. Convex clustering, in particular stood out by matching or even surpassing their performance.

In summary, convex clustering has attributes that make it valuable, in the realm of data clustering.

1. **Robustness:** Unlike techniques such as k-means, for clustering convex clustering demonstrates stability even when confronted with varying initial conditions. Nonetheless it remains necessary to select parameters in order to attain optimal performance.

2. **Flexibility:** Convex clustering is mainly used to identify clusters that have a convex-shape but it can also detect nonconvex patterns in specific situations. However, its effectiveness might differ depending on the characteristics of the dataset.
3. **Emphasis on Single Structure:** One notable aspect of convex clustering is its focus on generating cluster assignments. This emphasis proves advantageous when handling noisy data or datasets with gradual transitions between clusters, aiming to create clear and coherent cluster structures.

In applications, the choice of a clustering method should align with the inherent structure of the dataset, as there is no universally applicable solution for clustering tasks. In convex clustering, tuning the regularization parameter becomes crucial as it significantly influences the clustering outcomes.

However, achieving results in this case heavily relies on selecting the right parameters and aligning them with the specific characteristics of the dataset.

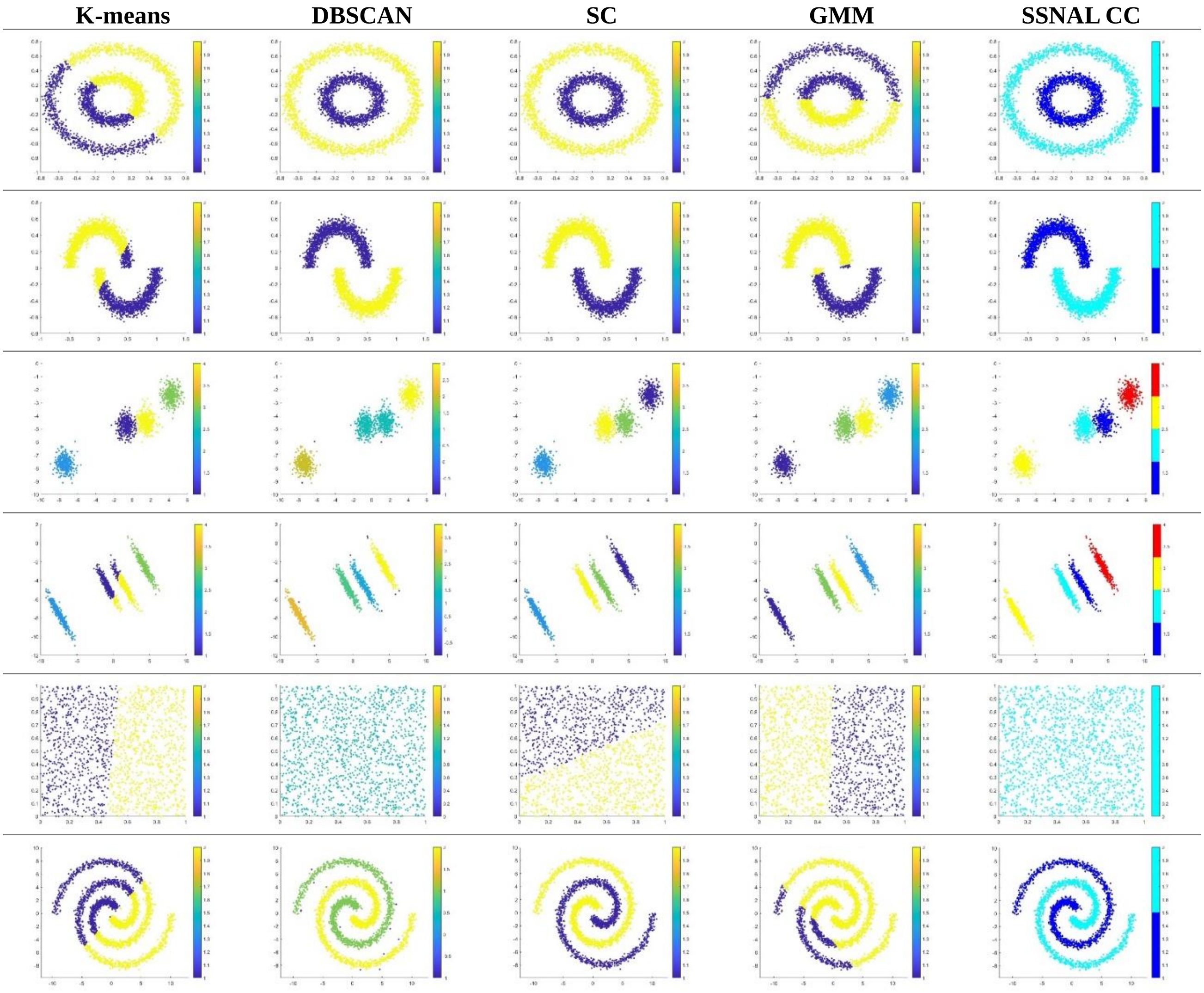


FIGURE 10. - A panoramic view of clustering techniques.

7. CHALLENGES AND FUTURE DIRECTIONS

As the field of optimization for clustering high-dimensional data continues to evolve researchers encounter both new challenges and opportunities for innovation. In this section we will explore the obstacles that researchers face and discuss potential directions, for future research.

7.1 ALGORITHMIC ENHANCEMENTS AND SCALABILITY

The potential for refining clustering algorithms is vast, especially with the rapid growth of high-dimensional data. Challenges such as computational complexity and sensitivity to parameters have become increasingly prominent. Okazaki and Kawano [114] observed a need for enhanced weight construction methods beyond conventional techniques like k-nearest neighbours, which could effectively address these challenges. Additionally, Li et al. [106] highlighted the challenges posed by local minima and suggested exploring algorithms that ensure global optima attainment or enhance robustness. There is also a pressing need for acceleration techniques, particularly for emerging clustering problems, as highlighted by Yuan et al. [99] and Shimmura et al. [101]. The scalability of clustering algorithms, especially in distributed environments, is another frontier yet to be fully explored, as indicated by Fodor et al. [102]. Lastly, investigating biconvex and nonconvex clustering methods, as discussed by Chakraborty and Xu [105], can provide greater flexibility in capturing intricate data structures.

7.2 VISUALIZATION, INTERPRETABILITY, AND ROBUSTNESS

Visualization and interpretability are crucial aspects of data analysis. However, the significant computational cost of convex clustering, as noted by Zhang et al. [90], poses a challenge, particularly for large-scale datasets. Addressing this challenge may involve the development of real-time dynamic visualization methods. Moreover, integrating sparse representations for denoising, as proposed by Weylandt et al. [91], can lead to clearer cluster formations. In addition to computational challenges, the increasing prevalence of noisy data demands clustering algorithms that are inherently robust against outliers, as emphasized by Quan and Chen [47]. Additionally, the automatic tuning of hyperparameters, which often requires careful calibration for optimal performance, remains an area with great potential for innovation.

7.3 APPLICATIONS, INTEGRATION, AND INTERDISCIPLINARY RESEARCH

Convex clustering techniques are increasingly being integrated into various disciplines, leading to novel interdisciplinary insights. In the field of medical and biological information learning, Yao and Allen [115] made significant advancements by introducing the clustered Gaussian graphical model (cGGM) in conjunction with a sophisticated symmetric convex clustering approach encapsulated within a singular, cohesive framework. This approach has proven to be highly effective, in identifying clusters linked to neurons. Complementing this, the potential of integrating biological prior knowledge into multi-omic data clustering, as highlighted by Shetta et al. [97], underscores the depth of insights that can be garnered through such sophisticated methods. Additionally, the realms of biclustering and tensor clustering, expanded upon by Chen and Suzuki [89] and Weylandt [78], present new avenues for innovative research. These methodologies are especially persistent given the complexity of higher-dimensional structures in data, which is a common challenge in biological data analysis.

In document clustering, convex clustering is a collecting of text data documents into groups of similar documents. This is important in high-dimension data, where efficient handling of large size of textual acquaintance is critical. In the field of image processing, the adaptability and effectiveness of convex clustering have been further underscored by recent research. Wang et al. [74] developed a novel, robust convex clustering method characterized by its resilience to withstand noise and outliers, thereby significantly improving the dependability of clustering results in image datasets. Furthermore, Condat [116] has proposed a novel convex formulation of the k-means algorithm, uniquely designed for both clustering and image segmentation purposes.

The interaction between convex clustering and contemporary machine learning approaches is quite interesting. This collaboration, especially when it comes to neural networks signifies a thrilling progress, in unsupervised feature acquisition paired with clustering. Consequently, it broadens the horizons of what can be accomplished through these methods.

Furthermore, the application of convex clustering in community detection in networks with potentially highly skewed degree distributions, is limited by the assumption that all nodes in the same community are statistically equivalent and have equal expected degrees [117]. In the financial industry, the adaptive convex clustering model with iteratively weighted least squares-based algorithm (ACC-IWLS) [118], which is effectively used to predict the likelihood of purchases for personalized requests. The application of convex clustering techniques in fields, such, as healthcare and finance demonstrate its potential to generate new and valuable interdisciplinary insights.

8. CONCLUSIONS

This detailed review targeting data clustering evolving. Particularly on applications robustness of convex clustering techniques. This focus mainly involves SSNAL algorithm. Convex clustering is recognized as a robust alternative to traditional clustering methods, amalgamating the strengths of conventional approaches and ensuring globally optimal solutions. The meticulous examination of synthetic datasets in relation to the SSNAL algorithm has enriched our understanding of the interplay between data and algorithmic techniques. These datasets often possess unique structures that challenge the algorithm's adaptability and underscore the importance of parameter sensitivity, particularly the gamma value, in cluster formation. Comparatively, convex clustering demonstrates superior performance over traditional methods like k-means and DBSCAN, especially in maintaining intra-cluster cohesion and clear inter-cluster separation, even in complex data structures. Furthermore, integrating convex clustering with modern machine learning paradigms, such as deep neural networks, opens promising avenues for future research. Despite challenges such as computational constraints in large datasets and noise management in high-dimensional spaces, the continuous evolution of algorithmic techniques and their interdisciplinary applications indicates that efficient, scalable and interpretable solutions are forthcoming.

Funding

None

ACKNOWLEDGEMENT

The first author would like to thank the reviewers for their invaluable feedback, which greatly improved the overall quality of this paper's presentation.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

REFERENCES

[1] G. C. Ngo and E. Q. B. Macabebe, "Image segmentation using K-means color quantization and density-based spatial clustering of applications with noise (DBSCAN) for hotspot detection in photovoltaic modules," in 2016 IEEE Region 10 Conference (TENCON), 2016, pp. 1614–1618. doi: 10.1109/TENCON.2016.7848290.

[2] G. Rudrappa, "Cloud Classification Using Ground Based Images Using CBIR and K-Means Clustering," Biosci Biotechnol Res Commun, vol. 13, no. 13, pp. 95–99, Dec. 2020, doi: 10.21786/bbrc/13.13/13.

[3] M. R. Karim et al., "Deep learning-based clustering approaches for bioinformatics," Brief Bioinform, vol. 22, no. 1, pp. 393–415, Jan. 2021, doi: 10.1093/bib/bbz170.

[4] M. Nascimento et al., "Independent Component Analysis (ICA) based-clustering of temporal RNA-seq data," PLoS One, vol. 12, no. 7, pp. e0181195, Jul. 2017, doi: 10.1371/journal.pone.0181195.

[5] S. A. Hussein and I. A. Zahid, "Improved Naked Mole-Rat Algorithm Based on Variable Neighborhood Search for the N-Queens Problem," Iraqi Journal of Science, 2024, doi: 10.24996/ijs.2024.65.1.41.

[6] S. A. Hussein and A. Y. Yousif, "An Improved Meerkat Clan Algorithm for Solving 0-1 Knapsack Problem," Iraqi Journal of Science, pp. 773–784, 2022, doi: 10.24996/ijs.2022.63.2.32.

[7] A. E. Ezugwu et al., "A comprehensive survey of clustering algorithms: State-of-the-art machine learning applications, taxonomy, challenges, and future research prospects," Eng Appl Artif Intell, vol. 110, p. 104743, 2022, doi:10.1016/j.engappai.2022.104743.

[8] A. Saxena et al., "A review of clustering techniques and developments," Neurocomputing, vol. 267, pp. 664–681, 2017, doi:10.1016/j.neucom.2017.06.053.

[9] N. H. Salman and S. N. Mohammed, "Image Segmentation Using PSO-Enhanced K-Means Clustering and Region Growing Algorithms," Iraqi Journal of Science, pp. 4988–4998, 2021, doi: 10.24996/ijs.2021.62.12.35.

[10] P. Fränti and S. Sieranoja, "K-means properties on six clustering benchmark datasets," Applied Intelligence, vol. 48, no. 12, pp. 4743–4759, 2018, doi: 10.1007/s10489-018-1238-7.

[11] F. Nie, J. Xue, D. Wu, R. Wang, H. Li, and X. Li, "Coordinate Descent Method for k-means," IEEE Trans Pattern Anal Mach Intell, vol. 44, no. 5, pp. 2371–2385, 2022, doi: 10.1109/TPAMI.2021.3085739.

[12] J. M. Pena, J. A. Lozano, and P. Larranaga, "An empirical comparison of four initialization methods for the k-means algorithm," Pattern Recognit Lett, vol. 20, no. 10, pp. 1027–1040, 1999, doi:10.1016/S0167-8655(99)00069-0.

[13] D. Arthur and S. Vassilvitskii, "k-means++: the advantages of careful seeding," in Proceedings of the Eighteenth Annual ACM-SIAM Symposium on Discrete Algorithms, in SODA '07. USA: Society for Industrial and Applied Mathematics, 2007, pp. 1027–1035.

[14] J. Peng and Y. Wei, "Approximating K‐means‐type Clustering via Semidefinite Programming," SIAM Journal on Optimization, vol. 18, no. 1, pp. 186‐205, Jan. 2007, doi: 10.1137/050641983.

[15] P. Fränti and O. Virmajoki, "Iterative shrinking method for clustering problems," Pattern Recognit, vol. 39, no. 5, pp. 761–775, 2006, doi: 10.1016/j.patcog.2005.09.012.

[16] M. Muhr and M. Granitzer, "Automatic Cluster Number Selection Using a Split and Merge K-Means Approach," in 2009 20th International Workshop on Database and Expert Systems Application, 2009, pp. 363–367. doi: 10.1109/DEXA.2009.39.

[17] X.-D. Wang, R.-C. Chen, Z.-Q. Zeng, C.-Q. Hong, and F. Yan, "Robust Dimension Reduction for Clustering With Local Adaptive Learning," IEEE Trans Neural Netw Learn Syst, vol. 30, no. 3, pp. 657–669, 2019, doi: 10.1109/TNNLS.2018.2850823.

[18] S. Iannucci, V. Cardellini, O. D. Barba, and I. Banicescu, "A hybrid model-free approach for the near-optimal intrusion response control of non-stationary systems," Future Generation Computer Systems, vol. 109, pp. 111–124, 2020, doi: 10.1016/j.future.2020.03.018.

[19] Á. Herrero, A. Jiménez, and S. Bayraktar, "Hybrid Unsupervised Exploratory Plots: A Case Study of Analysing Foreign Direct Investment," Complexity, vol. 2019, p. 6271017, 2019, doi: 10.1155/2019/6271017.

[20] F. Lindsten, H. Ohlsson, and L. Ljung, "Clustering using sum-of-norms regularization: With application to particle filter output computation," in 2011 IEEE Statistical Signal Processing Workshop (SSP), 2011, pp. 201–204. doi: 10.1109/SSP.2011.5967659.

[21] D. Sun, K.-C. Toh, and Y. Yuan, "Convex clustering: Model, theoretical guarantee and efficient algorithm," The Journal of Machine Learning Research, vol. 22, no. 1, pp. 427–458, 2021.

[22] M. Faizan, M. F. Zuhairi, S. Ismail, S. Sultan, and K. Lumpur, "Applications of Clustering Techniques in Data Mining: A Comparative Study," 2020. [Online]. Available: www.ijacsa.thesai.org

[23] A. K. Jain and R. C. Dubes, Algorithms for clustering data. Prentice-Hall, Inc., 1988.

[24] H. Petwal and R. Rani, "An Efficient Clustering Algorithm for Mixed Dataset of Postoperative Surgical Records," International Journal of Computational Intelligence Systems, vol. 13, no. 1, pp. 757–770, 2020, doi: 10.2991/ijcis.d.200601.001.

[25] S. Harikumar and S. PV, "K-Medoid Clustering for Heterogeneous DataSets," Procedia Comput Sci, vol. 70, pp. 226–237, 2015, doi: 10.1016/j.procs.2015.10.077

[26] V. Makara, M. V. Reddy, M. Vivekananda, and H. -Telangana, "Divisive Hierarchical Clustering with K-means and Agglomerative Hierarchical Clustering," International Journal of Computer Science Trends and Technology (IJCST), vol. 5, 2013, [Online]. Available: www.ijcstjournal.org

[27] H. C. Mandhare and S. R. Idate, "A comparative study of cluster based outlier detection, distance based outlier detection and density based outlier detection techniques," in 2017 International Conference on Intelligent Computing and Control Systems (ICICCS), 2017, pp. 931–935. doi: 10.1109/ICCONS.2017.8250601.

[28] B. A. Hussain and A. D. Salman, "Gene Expression Analysis via Spatial Clustering and Evaluation Indexing," Iraqi Journal for Computer Science and Mathematics, vol. 4, no. 1, pp. 24–34, 2023, doi: 10.52866/ijcsm.2023.01.01.004.

[29] S. K. Appiah, K. Wirekoh, E. N. Aidoo, S. D. Oduro, and Y. D. Arthur, "A model-based clustering of expectation–maximization and K-means algorithms in crime hotspot analysis," Research in Mathematics, vol. 9, no. 1, p. 2073662, Dec. 2022, doi: 10.1080/27684830.2022.2073662.

[30] Y. Zhang et al., "Gaussian mixture model clustering with incomplete data," ACM Transactions on Multimedia Computing, Communications and Applications, vol. 17, no. 1s, Apr. 2021, doi: 10.1145/3408318.

[31] U. von Luxburg, "A Tutorial on Spectral Clustering," Nov. 2007, [Online]. Available: http://arxiv.org/abs/0711.0189

[32] G. Li, "Generalized Co-clustering Analysis via Regularized Alternating Least Squares," Comput Stat Data Anal, vol. 150, p. 106989, 2020, doi: 10.1016/j.csda.2020.106989.

[33] L. Lazzeroni and A. Owen, "Plaid models for gene expression data," Stat Sin, vol. 12, no. 1, pp. 61–86, 2002, [Online]. Available: http://www.jstor.org/stable/24307036

[34] I. S. Dhillon, "Co-clustering documents and words using bipartite spectral graph partitioning," in Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining, 2001, pp. 269–274.

[35] B. A. Attea, "A fuzzy multi-objective particle swarm optimization for effective data clustering," Memet Comput, vol. 2, no. 4, pp. 305–312, 2010, doi: 10.1007/s12293-010-0047-2.

[36] S.-H. Lee, Y.-S. Jeong, J.-Y. Kim, and M. K. Jeong, "A new clustering validity index for arbitrary shape of clusters," Pattern Recognit Lett, vol. 112, pp. 263–269, 2018, doi: 10.1016/j.patrec.2018.08.005.

[37] X. Li, W. Liang, X. Zhang, S. Qing, and P.-C. Chang, "A cluster validity evaluation method for dynamically determining the near-optimal number of clusters," Soft comput, vol. 24, no. 12, pp. 9227–9241, 2020, doi: 10.1007/s00500-019-04449-7.

[38] K. Chowdhury, D. Chaudhuri, and A. K. Pal, "An entropy-based initialization method of K-means clustering on the optimal number of clusters," Neural Comput Appl, vol. 33, no. 12, pp. 6965–6982, 2021, doi: 10.1007/s00521-020-05471-9.

[39] J. Fang et al., "Clustering Deviation Index (CDI): a robust and accurate internal measure for evaluating scRNA-seq data clustering," Genome Biol, vol. 23, no. 1, p. 269, 2022, doi: 10.1186/s13059-022-02825-5.

[40] K. P. Sinaga and M.-S. Yang, "Unsupervised K-Means Clustering Algorithm," IEEE Access, vol. 8, pp. 80716–80727, 2020, doi: 10.1109/ACCESS.2020.2988796.

[41] Z. Safari, K. T. Mursi, and Y. Zhuang, "Fast Automatic Determination of Cluster Numbers for High Dimensional Big Data," in Proceedings of the 2020 4th International Conference on Compute and Data Analysis, in ICCDA '20. New York, NY, USA: Association for Computing Machinery, 2020, pp. 50–57. doi: 10.1145/3388142.3388164

[42] S. Yin, G. Gan, E. A. Valdez, and J. Vadiveloo, "Applications of clustering with mixed type data in life insurance," Risks, vol. 9, no. 3, pp. 1–19, Mar. 2021, doi: 10.3390/risks9030047.

[43] O. Al-Janabee and B. Al-Sarray, "Evaluation Algorithms Based on Fuzzy C-means for the Data Clustering of Cancer Gene Expression," Iraqi Journal for Computer Science and Mathematics, vol. 3, no. 2, pp. 27–41, 2022, doi: 10.52866/ijcsm.2022.02.01.004.

[44] T. Li, A. Rezaeipanah, and E. S. M. Tag El Din, "An ensemble agglomerative hierarchical clustering algorithm based on clusters clustering technique and the novel similarity measurement," Journal of King Saud University - Computer and Information Sciences, vol. 34, no. 6, 2022, doi: 10.1016/j.jksuci.2022.04.010.

[45] Z. Uykan, "Fusion of Centroid-Based Clustering With Graph Clustering: An Expectation-Maximization-Based Hybrid Clustering," IEEE Trans Neural Netw Learn Syst, vol. 34, no. 8, pp. 4068–4082, 2023, doi: 10.1109/TNNLS.2021.3121224.

[46] L. Song, Z. Ge, and E. Y. Lam, "Dual Alternating Direction Method of Multipliers for Inverse Imaging," IEEE Transactions on Image Processing, vol. 31, pp. 3295–3308, 2022, doi: 10.1109/TIP.2022.3167915.

[47] Z. Quan and S. Chen, "Robust convex clustering," Soft comput, vol. 24, no. 2, pp. 731–744, 2020, doi: 10.1007/s00500-019-04471-9.

[48] C. H. Nguyen and H. Mamitsuka, "On Convex Clustering Solutions," May 2021, [Online]. Available: http://arxiv.org/abs/2105.08348

[49] T. Liu, J. Zhu, J. Zhou, Y. Zhu, and X. Zhu, "Initialization-similarity clustering algorithm," Multimed Tools Appl, vol. 78, no. 23, pp. 33279–33296, 2019, doi: 10.1007/s11042-019-7663-8.

[50] S. Ma and J. Huang, "A Concave Pairwise Fusion Approach to Subgroup Analysis," J Am Stat Assoc, vol. 112, no. 517, pp. 410–423, Jan. 2017, doi: 10.1080/01621459.2016.1148039.

[51] M. Weylandt, J. Nagorski, and G. I. Allen, "Dynamic Visualization and Fast Computation for Convex Clustering via Algorithmic Regularization," Jan. 2019, doi: 10.1080/10618600.2019.1629943.

[52] B. Wang, Y. Zhang, W. W. Sun, and Y. Fang, "Sparse Convex Clustering," Journal of Computational and Graphical Statistics, vol. 27, no. 2, pp. 393–403, Apr. 2018, doi: 10.1080/10618600.2017.1377081.

[53] R. Suganya, M. Pavithra, and P. Nandhini, "Algorithms and Challenges in Big Data Clustering," International Journal of Engineering and Techniques, vol. 4, no. 4, pp. 40–47, 2018, [Online]. Available: http://www.ijetjournal.org

[54] J. Oyelade et al., "Clustering algorithms: Their application to gene expression data," Bioinform Biol Insights, vol. 10, pp. 237–253, Nov. 2016, doi: 10.4137/BBI.S38316.

[55] M. Vichi, C. Cavicchia, and P. J. F. Groenen, "Hierarchical Means Clustering," J Classif, vol. 39, no. 3, pp. 553–577, Nov. 2022, doi: 10.1007/s00357-022-09419-7.

[56] S. Weng, J. Gou, and Z. Fan, "h-DBSCAN: A simple fast DBSCAN algorithm for big data," in Asian Conference on Machine Learning, PMLR, 2021, pp. 81–96.

[57] K. Djouzi and K. Beghdad-Bey, "A Review of Clustering Algorithms for Big Data," in 2019 International Conference on Networking and Advanced Systems (ICNAS), 2019, pp. 1–6. doi: 10.1109/ICNAS.2019.8807822.

[58] Y. Lv, Z. Lan, C. Kan, and X. Zheng, "Polycentric Urban Development and its Determinants in China: A Geospatial Big Data Perspective," Geogr Anal, vol. 53, no. 3, pp. 520–542, Jul. 2021, doi: 10.1111/gean.12236.

[59] J. Blanza, "Wireless Propagation Multipaths using Spectral Clustering and Three-Constraint Affinity Matrix Spectral Clustering," Baghdad Science Journal, vol. 18, no. 2 (Suppl.), p. 1001, Jun. 2021, doi: 10.21123/bsj.2021.18.2(Suppl.).1001.

[60] H. Turner, T. Bailey, and W. Krzanowski, "Improved biclustering of microarray data demonstrated through systematic performance tests," Comput Stat Data Anal, vol. 48, no. 2, pp. 235–254, 2005, doi: 10.1016/j.csda.2004.02.003.

[61] T. Siswantining, A. Eriza Aminanto, D. Sarwinda, and O. Swasti, "Biclustering analysis using plaid model on gene expression data of colon cancer," Austrian Journal of Statistics, vol. 50, no. 5, pp. 101–114, 2021, doi: 10.17713/ajs.v50i5.1195.

[62] M. Tanveer, M. Tabish, and J. Jangir, "Sparse Pinball Twin Bounded Support Vector Clustering," IEEE Trans Comput Soc Syst, vol. 9, no. 6, pp. 1820–1829, 2022, doi: 10.1109/TCSS.2021.3122828.

[63] A. Panahi, D. Dubhashi, F. D. Johansson, and C. Bhattacharyya, "Clustering by sum of norms: Stochastic incremental algorithm, convergence and cluster recovery," in International conference on machine learning, PMLR, 2017, pp. 2769–2777.

[64] F. Lindsten, H. Ohlsson, and L. Ljung, "Just Relax and Come Clustering! : A Convexification of k-Means Clustering," 2011. [Online]. Available: https://api.semanticscholar.org/CorpusID:45329170

[65] Q. Feng, C. L. P. Chen, and L. Liu, "A Review of Convex Clustering From Multiple Perspectives: Models, Optimizations, Statistical Properties, Applications, and Connections," IEEE Trans Neural Netw Learn Syst, pp. 1–21, 2023, doi: 10.1109/TNNLS.2023.3276393.

[66] K. M. Tan and D. Witten, "Statistical properties of convex clustering," Electron J Stat, vol. 9, no. 2, pp. 2324–2347, Aug. 2015, doi: 10.1214/15-EJS1074.

[67] E. C. Chi, B. J. Gaines, W. W. Sun, H. Zhou, and J. Yang, "Provable convex co-clustering of tensors," Journal of Machine Learning Research, vol. 21, no. 214, pp. 1–58, 2020.

[68] T. D. Hocking, A. Joulin, F. Bach, and J.-P. Vert, "Clusterpath an algorithm for clustering using convex fusion penalties," in 28th international conference on machine learning, 2011, p. 1.

[69] B. Al Sarray, S. Chrétien, P. Clarkson, and G. Cottez, "Enhancing Prony's method by nuclear norm penalization and extension to missing data," Signal Image Video Process, vol. 11, no. 6, pp. 1089–1096, 2017, doi: 10.1007/s11760-017-1062-2.

[70] D. G. Mixon, S. Villar, and R. Ward, "Clustering subgaussian mixtures by semidefinite programming," Inf inference, vol. 6, no. 4, pp. 389–415, Dec. 2017, doi: 10.1093/imaiai/iax001.

[71] L. Han and Y. Zhang, "Reduction techniques for graph-based convex clustering," in Proceedings of the AAAI Conference on Artificial Intelligence, 2016.

[72] C. Lu, S. Yan, and Z. Lin, "Convex Sparse Spectral Clustering: Single-View to Multi-View," IEEE Transactions on Image Processing, vol. 25, no. 6, pp. 2833–2843, 2016, doi: 10.1109/TIP.2016.2553459

[73] S. Poddar and M. Jacob, "Convex clustering and recovery of partially observed data," in 2016 IEEE International Conference on Image Processing (ICIP), IEEE, 2016, pp. 3498–3502.

[74] Q. Wang, P. Gong, S. Chang, T. S. Huang, and J. Zhou, "Robust Convex Clustering Analysis," in 2016 IEEE 16th International Conference on Data Mining (ICDM), 2016, pp. 1263–1268. doi: 10.1109/ICDM.2016.0170.

[75] X. L. Sui, L. Xu, X. Qian, and T. Liu, "Convex clustering with metric learning," Pattern Recognit, vol. 81, pp. 575–584, 2018, doi: https://doi.org/10.1016/j.patcog.2018.04.019.

[76] H. Chen, L. Kong, and Y. Li, "A Novel Convex Clustering Method for High-Dimensional Data Using Semiproximal ADMM," Math Probl Eng, vol. 2020, p. 9216351, 2020, doi: 10.1155/2020/9216351.

[77] T. Jiang and S. Vavasis, "Certifying clusters from sum-of-norms clustering," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.11355

[78] M. Weylandt, "Splitting Methods For Convex Bi-Clustering And Co-Clustering," in 2019 IEEE Data Science Workshop (DSW), 2019, pp. 237–242. doi: 10.1109/DSW.2019.8755599.

[79] R. Tibshirani, "Regression Shrinkage and Selection via the Lasso," Journal of the Royal Statistical Society. Series B (Methodological), vol. 58, no. 1, pp. 267–288, 1996, [Online]. Available: http://www.jstor.org/stable/2346178

[80] K. Pelckmans, J. De Brabanter, J. A. K. Suykens, and B. De Moor, "Convex clustering shrinkage," in PASCAL workshop on statistics and optimization of clustering workshop, 2005.

[81] R. Tibshirani, M. Saunders, S. Rosset, J. Zhu, and K. Knight, "Sparsity and Smoothness Via the Fused Lasso," J R Stat Soc Series B Stat Methodol, vol. 67, no. 1, pp. 91–108, Feb. 2005, doi: 10.1111/j.1467-9868.2005.00490.x.

[82] S. Boyd, N. Parikh, E. Chu, B. Peleato, and J. Eckstein, "Distributed optimization and statistical learning via the alternating direction method of multipliers," Foundations and Trends® in Machine learning, vol. 3, no. 1, pp. 1–122, 2011. doi: 10.1561/2200000016.

[83] C. Zhu, H. Xu, C. Leng, and S. Yan, "Convex optimization procedure for clustering: Theoretical revisit," Adv Neural Inf Process Syst, vol. 27, 2014.

[84] E. C. Chi and K. Lange, "Splitting Methods for Convex Clustering," Journal of Computational and Graphical Statistics, vol. 24, no. 4, pp. 994–1013, Oct. 2015, doi: 10.1080/10618600.2014.948181.

[85] Y. Yuan, D. Sun, and K.-C. Toh, "An efficient semismooth Newton based algorithm for convex clustering," in International Conference on Machine Learning, PMLR, 2018, pp. 5718–5726.

[86] G. Xu, Y. Xia, and H. Ji, "Weighted total variation based convex clustering," Aug. 2018, [Online]. Available: http://arxiv.org/abs/1808.09144

[87] M. Zhang, "Forward-stagewise clustering: An algorithm for convex clustering," Pattern Recognit Lett, vol. 128, pp. 283–289, 2019, doi: 10.1016/j.patrec.2019.09.014.

[88] X. Zhou, C. Du, and X. Cai, "An Efficient Smoothing Proximal Gradient Algorithm for Convex Clustering," Jun. 2020, [Online]. Available: http://arxiv.org/abs/2006.12592

[89] J. Chen and J. Suzuki, "An efficient algorithm for convex biclustering," Mathematics, vol. 9, no. 23, Dec. 2021, doi: 10.3390/math9233021.

[90] B. Zhang, J. Chen, and Y. Terada, "Dynamic visualization for L1 fusion convex clustering in near-linear time," in Uncertainty in Artificial Intelligence, PMLR, 2021, pp. 515–524.

[91] M. Weylandt, T. M. Roddenberry, and G. I. Allen, "Simultaneous Grouping and Denoising via Sparse Convex Wavelet Clustering," in 2021 IEEE Data Science and Learning Workshop (DSLW), 2021, pp. 1–8. doi: 10.1109/DSLW51110.2021.9523413.

[92] Z. Wang, X. Liu, and Q. Li, "A Euclidean Distance Matrix Model for Convex Clustering," May 2021, [Online]. Available: http://arxiv.org/abs/2105.04947

[93] X. Wang, H. Wang, S. Wang, and J. Yuan, "Convex clustering method for compositional data via sparse group lasso," Neurocomputing, vol. 425, pp. 23–36, 2021, doi: 10.1016/j.neucom.2020.10.105.

[94] M. Wang and G. I. Allen, "Integrative generalized convex clustering optimization and feature selection for mixed multi-view data," The Journal of Machine Learning Research, vol. 22, no. 1, pp. 2498–2571, 2021.

[95] K. Shimamura and S. Kawano, "Bayesian sparse convex clustering via global-local shrinkage priors," Comput Stat, vol. 36, no. 4, pp. 2671–2699, 2021, doi: 10.1007/s00180-021-01101-7.

[96] J. Pi, H. Wang, and P. M. Pardalos, "A dual reformulation and solution framework for regularized convex clustering problems," Eur J Oper Res, vol. 290, no. 3, pp. 844–856, 2021, doi: 10.1016/j.ejor.2020.09.010.

[97] O. Shetta, M. Niranjan, and S. Dasmahapatra, "Convex Multi-View Clustering Via Robust Low Rank Approximation With Application to Multi-Omic Data," IEEE/ACM Trans Comput Biol Bioinform, vol. 19, no. 6, pp. 3340–3352, 2022, doi: 10.1109/TCBB.2021.3122961.

[98] Y. Chen, C. Lei, C. Li, H. Ma, and N. Hu, "Robust convex biclustering with a tuning-free method," Dec. 2022, [Online]. Available: http://arxiv.org/abs/2212.03122

[99] Y. Yuan, T.-H. Chang, D. Sun, and K.-C. Toh, "A Dimension Reduction Technique for Large-Scale Structured Sparse Optimization Problems with Application to Convex Clustering," SIAM Journal on Optimization, vol. 32, no. 3, pp. 2294–2318, Sep. 2022, doi: 10.1137/21M1441080.

[100] D. J. W. Touw, P. J. F. Groenen, and Y. Terada, "Convex Clustering through MM: An Efficient Algorithm to Perform Hierarchical Clustering," Nov. 2022, [Online]. Available: http://arxiv.org/abs/2211.01877

[101] R. Shimmura and J. Suzuki, "Converting ADMM to a proximal gradient for efficient sparse estimation," Japanese Journal of Statistics and Data Science, vol. 5, no. 2, pp. 725–745, 2022, doi: 10.1007/s42081-022-00150-6.

[102] L. Fodor, D. Jakovetić, D. Boberić Krstićev, and S. Škrbić, "A parallel ADMM-based convex clustering method," EURASIP J Adv Signal Process, vol. 2022, no. 1, p. 108, 2022, doi: 10.1186/s13634-022-00942-8.

[103] A. Armacki, D. Bajovic, D. Jakovetic, and S. Kar, "Personalized Federated Learning via Convex Clustering," in 2022 IEEE International Smart Cities Conference (ISC2), 2022, pp. 1–7. doi: 10.1109/ISC255366.2022.9921863.

[104] B. Wang, L. Yao, J. Hu, and H. Li, "A New Algorithm for Convex Biclustering and Its Extension to the Compositional Data," Stat Biosci, vol. 15, no. 1, pp. 193–216, 2023, doi: 10.1007/s12561-022-09356-4.

[105] S. Chakraborty and J. Xu, "Biconvex Clustering," Journal of Computational and Graphical Statistics, vol. 32, no. 4, pp. 1524–1536, Oct. 2023, doi: 10.1080/10618600.2023.2197474.

[106] D. Li, S. Zhou, T. Zeng, and R. H. Chan, "Multi-Prototypes Convex Merging Based K-Means Clustering Algorithm," Feb. 2023, [Online]. Available: http://arxiv.org/abs/2302.07045

[107] I. T. Jollife and J. Cadima, "Principal component analysis: A review and recent developments," Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences, vol. 374, no. 2065. Royal Society of London, Apr. 13, 2016. doi: 10.1098/rsta.2015.0202.

[108] L. Van der Maaten and G. Hinton, "Visualizing data using t-SNE.," Journal of machine learning research, vol. 9, no. 11, 2008.

[109] S. Chrétien, C. Dombry, and A. Faivre, "The Guedon-Vershynin Semi-definite Programming Approach to Low Dimensional Embedding for Unsupervised Clustering," Front Appl Math Stat, vol. 5, Oct. 2019, doi: 10.3389/fams.2019.00041.

[110] J. K. Parker and L. O. Hall, "Accelerating Fuzzy-C Means Using an Estimated Subsample Size," IEEE Transactions on Fuzzy Systems, vol. 22, no. 5, pp. 1229–1244, 2014, doi: 10.1109/TFUZZ.2013.2286993.

[111] J.-P. Mei, Y. Wang, L. Chen, and C. Miao, "Large Scale Document Categorization With Fuzzy Clustering," IEEE Transactions on Fuzzy Systems, vol. 25, no. 5, pp. 1239–1251, 2017, doi: 10.1109/TFUZZ.2016.2604009.

[112] L. Hubert and P. Arabie, "Comparing partitions," J Classif, vol. 2, pp. 193–218, 1985.

[113] S. Modak, "Finding groups in data: an introduction to cluster analysis," J Appl Stat, pp. 1–3, doi: 10.1080/02664763.2023.2220087.

[114] A. Okazaki and S. Kawano, "Multi-Task Learning Regression via Convex Clustering," Apr. 2023, [Online]. Available: http://arxiv.org/abs/2304.13342

[115] T. Yao and G. I. Allen, "Clustered Gaussian Graphical Model Via Symmetric Convex Clustering," in 2019 IEEE Data Science Workshop (DSW), 2019, pp. 76–82. doi: 10.1109/DSW.2019.8755774.

[116] L. Condat, "A convex approach to K-means clustering and image segmentation," in Energy Minimization Methods in Computer Vision and Pattern Recognition: 11th International Conference, EMMCVPR 2017, Venice, Italy, October 30–November 1, 2017, Revised Selected Papers 11, Springer, 2018, pp. 220–234.

[117] Y. Chen, L. I. Xiaodong, and J. Xu, "Convexified modularity maximization for degree-corrected stochastic block models," Ann Stat, vol. 46, no. 4, pp. 1573–1602, Aug. 2018, doi: 10.1214/17-AOS1595.

[118] S. Chu, H. Jiang, Z. Xue, and X. Deng, "Adaptive Convex Clustering of Generalized Linear Models With Application in Purchase Likelihood Prediction," Technometrics, vol. 63, no. 2, pp. 171–183, Apr. 2021, doi: 10.1080/00401706.2020.1733094.