informatics

MDPI

Review

Clustering with Uncertainty: A Literature Review to Address a Cross-Domain Perspective †

Salvatore Flavio Pileggi

Faculty of Engineering and IT, University of Technology Sydney, P.O. Box 123, Ultimo, NSW 2007, Australia; salvatoreflavio.pileggi@uts.edu.au

† This paper is an extended version of our paper published in *Computational Science—ICCS 2024, Lecture Notes in Computer Science, vol 14838, Springer, Cham.* https://doi.org/10.1007/978-3-031-63783-4_22.

Abstract: Clustering is a very popular computational technique that, because of imperfect data, is often applied in the presence of some kind of uncertainty. Taking into account such an uncertainty (and model), the computational output accordingly contributes to increasing the accuracy of the computations and their effectiveness in context. However, there are challenges. This paper presents a literature review on the topic. It aims to identify and discuss the associated body of knowledge according to a cross-domain perspective. A semi-systematic methodology has allowed for the selection of 68 papers, prioritizing the most recent contributions and an intrinsic application-oriented approach. The analysis has underscored the relevance of the topic in the last two decades, in which computation has become somewhat pervasive in the context of inherent data complexity. Furthermore, it has identified a trend of domain-specific solutions over generic-purpose approaches. On one side, this trend enables a more specific set of solutions within specific communities; on the other side, the resulting distributed approach is not always well integrated with the mainstream. The latter aspect may generate a further fragmentation of the body of knowledge, mostly because of some lack of abstraction in the definition of specific problems. While in general terms these gaps are largely understandable within the research community, a lack of implementations to provide ready-to-use resources is critical overall. In more technical terms, solutions in the literature present a certain inclination to mixed methods, in addition to the classic application of Fuzzy Logic and other probabilistic approaches. Last but not least, the propagation of the uncertainty in the current technological context, characterised by data and computational intensive solutions, is not fully analysed and critically discussed in the literature. The conducted analysis intrinsically suggests consolidation and enhanced operationalization though Open Software, which is crucial to establish scientifically sound computational frameworks.

Keywords: clustering; uncertainty modelling; uncertainty management; unsupervised learning; data analysis; data mining

check for updates

Academic Editor: Olga Kurasova

Received: 16 January 2025
Revised: 1 April 2025
Accepted: 2 April 2025
Published: 9 April 2025

Citation: Pileggi, S.F. Clustering with Uncertainty: A Literature Review to Address a Cross-Domain Perspective. *Informatics* **2025**, *12*, 38. https://doi.org/10.3390/informatics12020038

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Empirical observations show an increasing quantity of data with a degree of uncertainty [1]. Indeed, real-world data naturally tend to present uncertainty due to different factors including, among others, human or instrumental errors [2], randomness, imprecision, vagueness, and partial ignorance [3]. In general terms, the theoretical impact of data uncertainty, as well as the risk associated with ignoring it (e.g., [4,5]), is a well-known issue within the scientific community and is, indeed, largely addressed in the literature. In general, it is strongly suggested that, wherever possible, a proper and

explicit uncertainty model should always be used to effectively support representation, visualization [6], measuring/quantifying, and consequent analysis. From a more practical perspective, more and more studies present a specific focus on uncertainty in a variety of application domains, such as, among the many methods, budget impact analysis [7], organizational environments [8] and hydrological data [9]. Such critical modelling is intrinsically challenging and may require a domain-specific approach, such as Big Data [10,11], visualization [12] and Deep Learning [13].

On the other side, clustering techniques [14] group data points into different categories (clusters) based on their similarity, computed according to a given formal metric. These techniques have been extensively used in a general scientific context and traditional approaches keep evolving as a response to an environment characterised by evolving needs [15]. For instance, clustering is a common class of unsupervised learning [16], often adapted to achieve concrete goals in the different application domains (e.g., [17]), and methods such as formal classification [18], ontological modelling [19,20], and rule mining [21] commonly rely on clustering techniques.

Intuitively, clustering in a context of uncertainty, or even just potential uncertainty, proposes additional significant challenges on both (i) modelling similarity between uncertain objects and (ii) developing effective and efficient computational methods accordingly [22]. Alternative approaches to deal with uncertainty can be used for different reasons in different contexts. A classification of these techniques is not trivial. For instance, in [23], the authors have identified two main broad categories that aim, respectively, to complement and to generalise probabilistic representations. The former family addresses non-probabilistic uncertainty (typically imprecision, vagueness, or gradedness), while the latter targets the effective modelling of partial ignorance. More holistically, looking at extensions of traditional methods, three main categories have been summarised in [22]: partitioning clustering, density-based clustering, and possible world approaches. The resulting extended solutions integrate the original semantics with uncertainty modelling.

In continuity with an established body of knowledge, this papers aims to holistically review the most recent contributions and related advances in the field of clustering in the presence of uncertainty. Indeed, computational frameworks are continuously evolving as a response to emerging applications and changing requirements. Such a process is a driver for novel solutions, such as Deep Embedded Clustering [24] and Graph Neuronal Networks [25], as well as a determinant to adapt and apply existing techniques and methods. The explicit application focus further contributes to consolidate the body of knowledge in the field according to a cross-domain perspective resulting from a contextual analysis performed in the context of the most relevant computational trends and related applications.

1.1. Related Work

This work can be framed in the very broad context of uncertain data algorithms and applications [26]. A valuable review specifically on clustering has been provided in 2017 [3]. The focus of such a work is on uncertainty management and associated theoretical formalisms. Other concise contributions aiming at summarising the body of knowledge (e.g., [1,23]) are definitely valuable but also relatively dated, given the strong and constant advances in the computational world.

This paper provides an additional contribution to the body of knowledge in the field by addressing an application perspective with a focus on recent advances. Such an approach minimises the overlapping with existing reviews by integrating the already consolidated theoretical foundations with an application-oriented analysis. As explained later on in the paper (Section 3), the analysis framework has been designed accordingly.

1.2. Previous Work

A preliminary version of the paper has been published in 2024 at the International Conference in Computational Science (ICCS 2024) [27]. The conference version proposes a concise, yet self-contained, analysis that has been extended to provide a more exhaustive contribution by:

- Explanations have been enhanced and a conceptualization (Section 2 has been added, and most parts of the paper have been broadened accordingly).
- Extending the critical analysis and discussion in Sections 4 and 5.
- Generic improvement, as the paper has been holistically revised in all its parts.

These extensions provide a clear added value to the original work from both a conceptual and a critical point of view.

1.3. Structure of the Paper

The introductory part of the paper follows with a more extensive conceptual description of clustering in presence of uncertainty (Section 2) and is concluded by a description of the adopted methodology (Section 3). The core part of the paper includes two different sections that aim, respectively, to overview the most relevant contributions in the literature according to a cross-domain perspective (Section 4) and to discuss the results by looking at major gaps and challenges (Section 5). Finally, Section 6 provides an overview of the work.

2. Clustering with Uncertainty

Clustering is an intuitive concept that aims to partition a given data space by grouping data objects into different clusters according to their characteristics. Clustering techniques typically identify existing similarities among data points to provide consistent classification.

From a Knowledge Engineering perspective, this data classification from a uniform data space may be considered crucial, if not determinant, to establish semantics and, more in general, to support the application of advanced computational systems. The conceptual relevance of clustering has been historically recognised (e.g., [28]), as well as its practical relevance across the different disciplines (e.g., [29]).

A simplified overview of the clustering concept (Figure 1) assumes two main generic driving factors that determine the result: the method, understood as the technique to cluster data, and the number of clusters to consider. The former defines the similarity criteria, while the latter is normally estimated heuristically (e.g., by adopting the elbow method [30]). A relatively recent work [31] proposes a consistent taxonomy of clustering approaches. It distinguishes between two main classes (hierarchical and partitional) and defines a number of sub-classes accordingly.

Depending on the nature of the considered data (e.g., dimensionality), its distribution in the space and the adopted clustering technique, we may observe considerably different areas of potential overlapping (Figure 2). Such areas normally include data objects that are similar with characteristics that are borderline, meaning that their association with one cluster or another is determined by small values of the considered metrics. It intrinsically suggests a probabilistic approach over a more traditional Boolean logic to associate a data point to a given cluster. Indeed, probabilistic models potentially allow for uncertainty identification, as well as its quantification and consequent incorporation as part of the algorithm output. For applications that are sensitive or critical in nature (such as, among many, medical diagnostics, cybersecurity, and decision making), the clustering-induced uncertainty may be as impactful as the more explicit underlying data uncertainty.

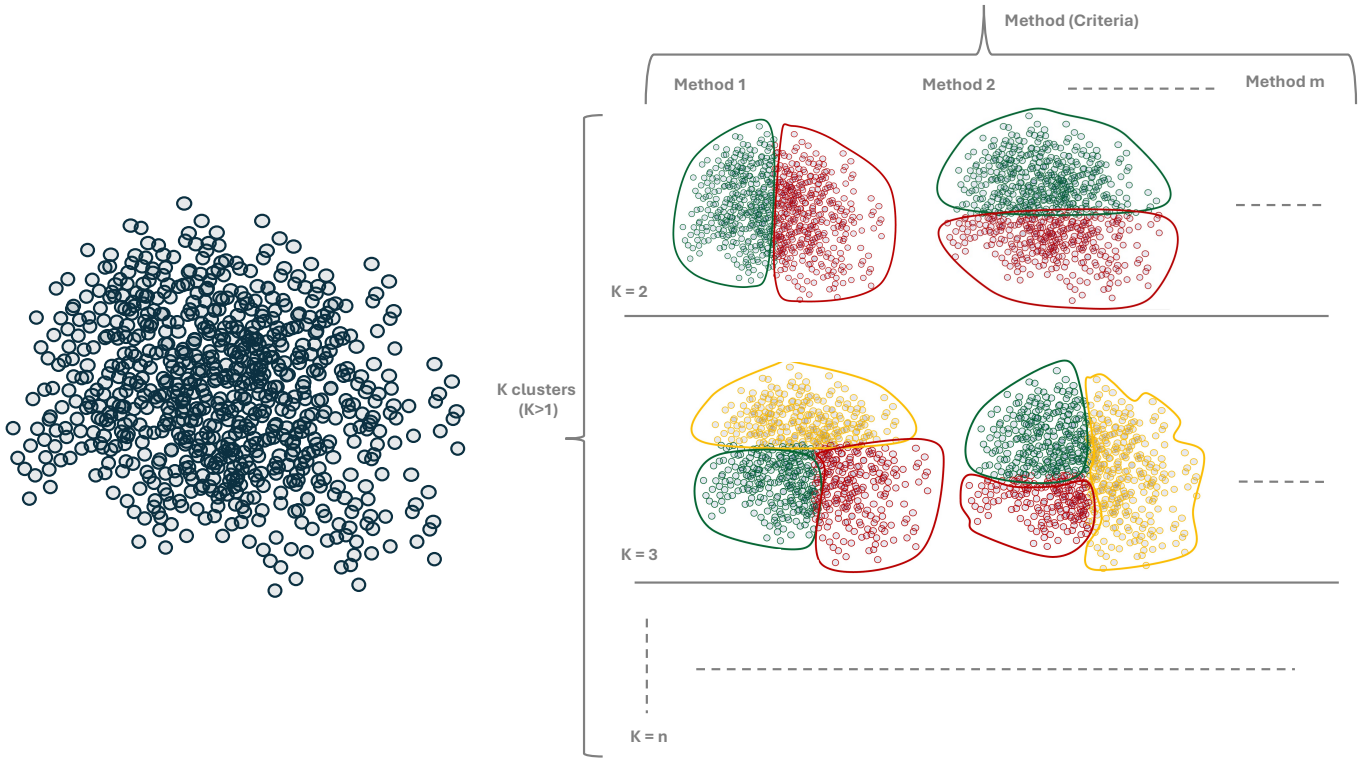


Figure 1. A simplified view of clustering.

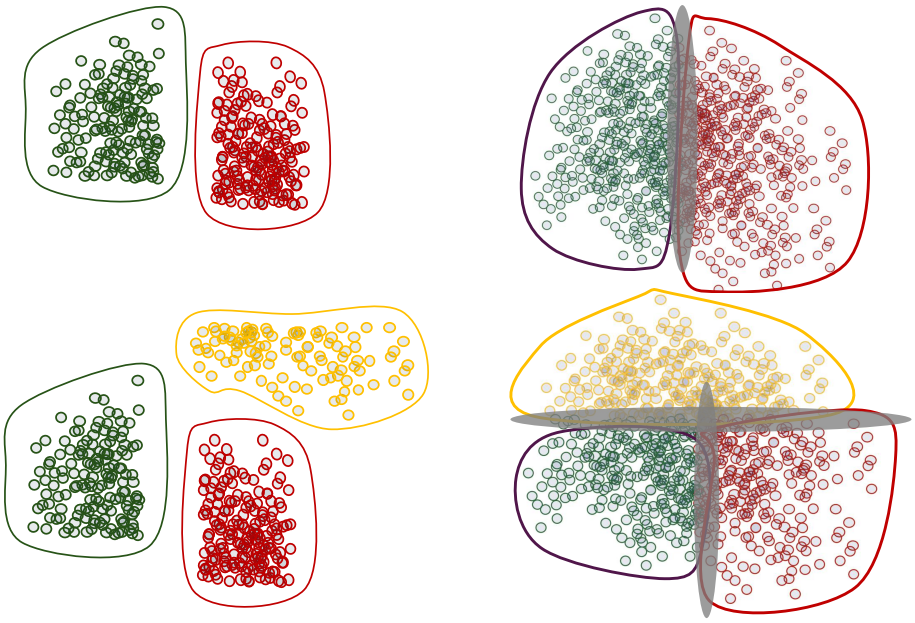


Figure 2. Areas of potential overlapping in the partition.

An explicit uncertainty is introduced by imperfect data, as discussed early on in the paper. As shown in Figure 3, such an imperfection is related to a probabilistic representation of data. Indeed, while "perfect" data are uniquely associated with a point in the data space, the representation of imperfect data is associated with a probabilistic pattern. Given a uncertainty in the input data, the outcome of clustering is expected to reflect that uncertainty.

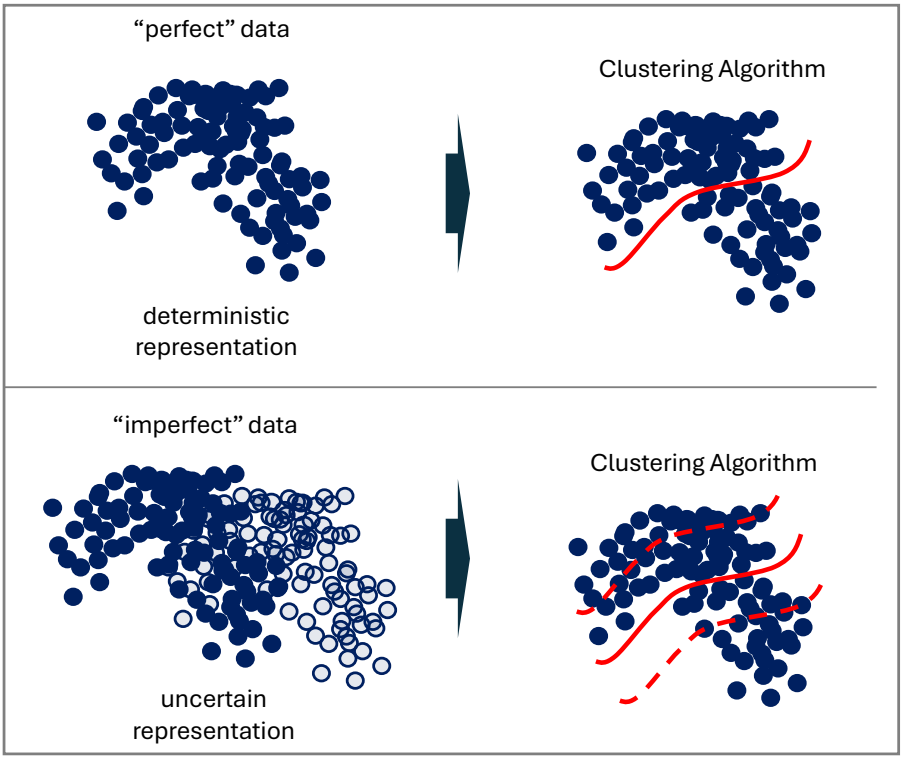


Figure 3. From “perfect” data to uncertainty.

In traditional clustering, a given data point is assigned to a single cluster according to Boolean logic. Clustering with uncertainty adopts a probabilistic logic instead, normally assuming that data points can potentially belong to multiple clusters. The most popular class of solutions adopts Fuzzy Logic [32], while other approaches for Uncertainty Quantification (UQ) in the specific field of clustering are based, among others, on Variation Inference for approximating complex probability densities [33], Deep Embedded Clustering [24] and related variants, and Graph Neuronal Networks [25]. However, because of the foci of this work, which prioritizes an application perspective, an explicit emphasis on emerging solutions may be limited.

3. Methodology and Approach

In order to generate a tangible contribution to the body of knowledge and avoid, as much as possible, overlapping and a lack of depth, this literature review has been conducted by combining a typical systematic process with non-systematic practices. The latter have been considered to address a focused search in a context where terminology may present a significant diversity. Indeed, while the very generic keywords adopted to retrieve papers from the different databases cannot assure comprehensiveness, snowballing from related references enables broader exploration capabilities. Additionally, the relatively soft inclusion criteria to properly address a cross-domain perspective intrinsically reduce the systematic character of the method.

The mainstream process assumes, as usual, paper retrieval from relevant databases. In this specific case, queries have been performed by simply combining two main keywords, namely *Clustering* and *Uncertainty*, to put emphasis on contributions that explicitly deal with uncertainty. Overall, the adopted methodology reflects an attempt to capture and critically re-elaborate an application perspective, rather than re-proposing the typical algorithm-

based analysis. The latter is extremely interesting but also already largely addressed in the literature.

3.1. Selection Criteria and Saturation

The selection of the papers to include in the review has been performed by applying a critical analysis aimed at the identification of the most relevant contributions in the field. The aimed application-oriented analysis inherently suggests a focus on "modern" systems. However, because of the objective difficulty to translate such an abstracted concept into a predefined time-range, a preliminary scanning has been performed. Looking at the scale and complexity of the different systems, as well as at the related technological evolution, this preliminary phase has suggested a focus on the last two decades (2005 onward). Such a time-frame seems effective to narrow the search, highlight the most recent advances in context, and maximise the provided value. The relatively soft selection criteria enabled the retrieval of an important number of papers. However, the selection process has in fact been much more focused. Indeed, after a number of iterations, a feeling of saturation naturally emerged as contributions started to present consolidations of existing concepts rather than novel solutions. This additional non-systematic element has been a determinant to facilitate de facto conciseness at a relevant scale.

3.2. Analysis Framework and Limitations

The analysis has been conducted according to two major dimensions: *domain* and *approach*. The former dimension aims primarily to distinguish between generic-purpose and domain-specific solutions, while the latter wants to facilitate an overview of major techniques. The presentation of the review (Section 4) has been structured by looking at the domain. Indeed, the classification of the different approaches is intrinsically more fragmented and not always explicit. In general terms, the classification followed the claims by authors and the original analysis. Non-systematic practices may have introduced biases. This applies mostly to selection criteria. Additionally, because of the high number of existing works distributed in a variety of domains, it is hard to assess the exhaustiveness of the review. Last but not least, no qualitative analysis has been conducted to minimise subjective assessments given the objective difficulty to identify reasonable criteria consistent with this specific case study.

4. A Cross-Domain Analysis

This section has a descriptive focus as it provides an overview of the contributions included in this study. It is structured to reflect a double perspective, including an approach and a domain analysis. The former (Section 4.1) focuses on the underlying methods, while the latter (Section 4.4) overviews the application domain and, indirectly, the related sources of uncertainty.

4.1. Approach Analysis

We deal separately with solutions that present a completely generic focus (referred to as "generic-purpose" and reported in Table 1) and those that have been designed within a specific application domain ("domain-specific", Table 2). This generic classification naturally introduces a cross-domain analysis. However, there are not always well-defined boundaries, as certain applications as identified in the context of this work may present a certain degree of genericness.

Table 1. Generic-purpose selected contributions.

Title/Ref.	Year	Approach
Cloud-Cluster: An uncertainty clustering algorithm based on cloud model [34]	2023	Method
Outlier-robust multi-view clustering for uncertain data [35]	2021	Multi-view Clustering
Multi-view spectral clustering for uncertain objects [36]	2021	Multi-view Clustering
Modeling uncertain data using Monte Carlo integration method for clustering [37]	2019	Monte-Carlo
Optimal clustering under uncertainty [38]	2018	Method
Locally weighted ensemble clustering [39]	2017	Ensemble Clustering
Self-adapted mixture distance measure for clustering uncertain data [40]	2017	Method
Novel density-based and hierarchical density-based clustering algorithms for uncertain data [41]	2017	Hierarchical Clustering
An information-theoretic approach to hierarchical clustering of uncertain data [42]	2017	Hierarchical Clustering
Active Clustering with Model-Based Uncertainty Reduction [43]	2016	Active Clustering
Robust ensemble clustering using probability trajectories [44]	2015	Ensemble Clustering
Large margin clustering on uncertain data by considering probability distribution similarity [45]	2015	PD Similarity
Representative clustering of uncertain data [46]	2014	Framework
Active spectral clustering via iterative uncertainty reduction [47]	2012	Active Clustering
Minimizing the variance of cluster mixture models for clustering uncertain objects [48]	2012	Method
Clustering uncertain data based on probability distribution similarity [22]	2011	PD Similarity
Clustering uncertain data using voronoi diagrams and r-tree index [49]	2010	Method
Subspace clustering for uncertain data [50]	2010	Sub-space clustering
Exceeding expectations and clustering uncertain data [51]	2009	Optimization
Clustering Uncertain Data with Possible Worlds [52]	2009	Method
Clustering Uncertain Data Via K-Medoids [53]	2008	Method
A hierarchical algorithm for clustering uncertain data via an information-theoretic approach [54]	2008	Hierarchical Clustering
Clustering Uncertain Data Using Voronoi Diagrams [55]	2008	Voronoi diagrams
Uncertain fuzzy clustering: Insights and recommendations [56]	2007	Fuzzy Logic
Uncertain fuzzy clustering: Interval type-2 fuzzy approach to c-means [57]	2007	Fuzzy Logic
Density-based clustering of uncertain data [58]	2005	Fuzzy Logic
Hierarchical density-based clustering of uncertain data [59]	2005	Fuzzy Logic

Table 2. Domain-specific selected contributions.

Title/Ref.	Year	Approach	Domain
Stochastic economic dispatch of wind power under uncertainty using clustering-based extreme scenarios [60]	2024	Stochastic Model	Energy
Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning [61]	2022	Fuzzy Logic	Machine Learning
A three-stage automated modal identification framework for bridge parameters based on frequency uncertainty and density clustering [62]	2022	Method	Engineering
Clustering uncertain graphs using ant colony optimization (ACO) [63]	2022	Optimization	Graphs
Uncertainty-Aware Clustering for Unsupervised Domain Adaptive Object Re-Identification [64]	2022	Hierarchical Clustering	Machine Learning
Decision-based scenario clustering for decision-making under uncertainty [65]	2022	Method	Decision Making
Active domain adaptation via clustering uncertainty-weighted embeddings [66]	2021	Active Learning	Machine Learning
UAC: An Uncertainty-Aware Face Clustering Algorithm [67]	2021	Method	Face Recognition
Uncertainty assessment in reservoir performance prediction using a two-stage clustering approach: Proof of concept and field application [68]	2021	Method	Petroleum Science
Handling uncertainty in financial decision making: a clustering estimation of distribution algorithm with simplified simulation [69]	2020	Method	Decision Making
Ride-sharing under travel time uncertainty: Robust optimization and clustering approaches [70]	2020	Method	Transportation
Deep semantic clustering by partition confidence maximisation [71]	2020	Method	Machine Learning
Uncertainty mode selection in categorical clustering using the rough set theory [72]	2020	Rough Set	Categorical Data
Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding [73]	2020	Fuzzy Logic	Energy
Efficient Assessment of Reservoir Uncertainty Using Distance-Based Clustering: A Review [74]	2019	Review	Petroleum Science
Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets [75]	2019	Fuzzy Logic	Genetics
Clustering mining blocks in presence of geological uncertainty [76]	2019	Method	Geology
Efficient and effective algorithms for clustering uncertain graphs [77]	2019	Optimization	Graphs
A three-way clustering approach for handling missing data using GTRS [78]	2018	Three-way Clustering	Missing Data
Clustering uncertain graphs [79]	2017	Method	Graphs
Novel adaptive multi-clustering algorithm-based optimal ESS sizing in ship power system considering uncertainty [80]	2017	Optimization	Energy
Multiple clustering views from multiple uncertain experts [81]	2017	Bayesian Model	Collaborative Environments
Uncertain data clustering in distributed peer-to-peer networks [82]	2017	Distributed Clustering	P2P Network
Efficient clustering of large uncertain graphs using neighborhood information [83]	2017	Framework	Graphs
Clustering based unit commitment with wind power uncertainty [84]	2016	Method	Energy
A framework for clustering uncertain data [85]	2015	Framework	Visualization

Table 2. *Cont.*

Title/Ref.	Year	Approach	Domain
Efficient clustering of uncertain data streams [86]	2014	Method	Data Stream
Uncertain data clustering-based distance estimation in wireless sensor networks [87]	2014	Method	Wireless Sensor Network
Weighted graph clustering with non-uniform uncertainties [88]	2014	Optimization	Graphs
Clustering large data with uncertainty [89]	2013	Fuzzy Logic	Large Data
Reliable clustering on uncertain graphs [90]	2012	Possible Worlds	Graphs
Clustering uncertain trajectories [91]	2011	Method	Location Data
Hue-stream: Evolution-based clustering technique for heterogeneous data streams with uncertainty [92]	2011	Method Data Stream	
An algorithm for clustering heterogeneous data streams with uncertainty [93]	2010	Method	Data Stream
On high dimensional projected clustering of uncertain data streams [94]	2009	Method	Data Stream
Clustering trajectories of moving objects in an uncertain world [95]	2009	Method	Location Data
A Framework for Clustering Uncertain Data Streams [96]	2008	Method	Data Stream
Conceptual clustering categorical data with uncertainty [97]	2007	Quality Assessment	Categorical Data
Including probe-level uncertainty in model-based gene expression clustering [98]	2007	Method	Genetics
Pvclust: an r package for assessing the uncertainty in hierarchical clustering [99]	2006	Hierarchical Clustering	Genetics
Uncertain Data Mining: An Example in Clustering Location Data [100]	2006	UK-Means	Location Data

4.2. Generic-Purpose Solutions

Among the generic-purpose works, there are two clearly identifiable sub-sets of solutions adopting, respectively, mixed (or not uniquely classifiable) methods [34,38,40,48,49,52,53] and Fuzzy Logic [56–59]. This is somehow in line with the key concepts previously introduced in the paper, as the application of Fuzzy Logic is probably the most common approach to deal with uncertainty in this specific context. On the other side, hybrid approaches are largely understandable in a context of inherent complexity.

Smaller classes of solutions adopt different clustering techniques to deal with uncertainty: Hierarchical Clustering [41,42,54], which builds a hierarchy of clusters [101]; Ensemble Clustering [39,44], which is based on the concept of "consensus" [102]; Multi-view Clustering [35,36], which explicitly overcomes the traditional single view for clustering [103]; and Active Clustering [43,47].

The probabilistic approach is relatively popular [22,45], while other methods are based on different approaches, including framework-based solutions [46], Voronoi diagrams [55], Monte Carlo [37], optimization strategies [51] and Sub-space Clustering [50].

4.3. Domain-Specific Solutions

Mixed methods [62,65,67–71,76,79,84,86,87,91–96,98], as well as Fuzzy Logic [61,73,75,89], Hierarchical Clustering [64,99], Optimization [63,77,80,88], and framework-based approaches [83,85] play a significant role also in a context of domain-specific applications. Other contributions include the adaptation of traditional techniques [100], quality assessment [97], Possible Worlds [90], Distributed Clustering [82], Bayesian Modeling [81], Three-way Clustering [78], a review for assessment purposes within a specific domain [74], rough set theory [72], Active Learning [66] and Stochastic Models [60].

4.4. Domain Analysis

As reported in Table 2, in the original contributions uncertainty is often related to computational techniques. That is in line with the modern popularity of soft computing, which normally assumes imprecision, uncertainty, partial truth, and approximations [104] and explains the notable focus on machine learning, graphs, and data streams.

Similarly, an alternative perspective of analysis associates the uncertainty with the characteristics of data. Apart from the already mentioned data streams, examples included in this review are a large dataset, location, and categorical data.

Looking at the applications, the review is characterised by a certain fragmentation with an emphasis on big generic domains.

5. Discussion

In order to critically discuss the review, the next section is structured in different subsections to address an overview of the results (Section 5.1), followed by a critical analysis of the major gaps emerging (Section 5.2) and, finally, more holistic considerations (Sections 5.3 and 5.4).

5.1. Overview

In quantitative terms, the majority (62%) of the 68 papers selected within the time range 2005–2024 are journal articles. A similar percentage (60%) presents a domain-specific focus. As shown in Figure 4, such a trend becomes more consistent and somehow predominant from 2018 onward. More holistically, the study confirms a substantial research interest in the topic throughout the observation period.

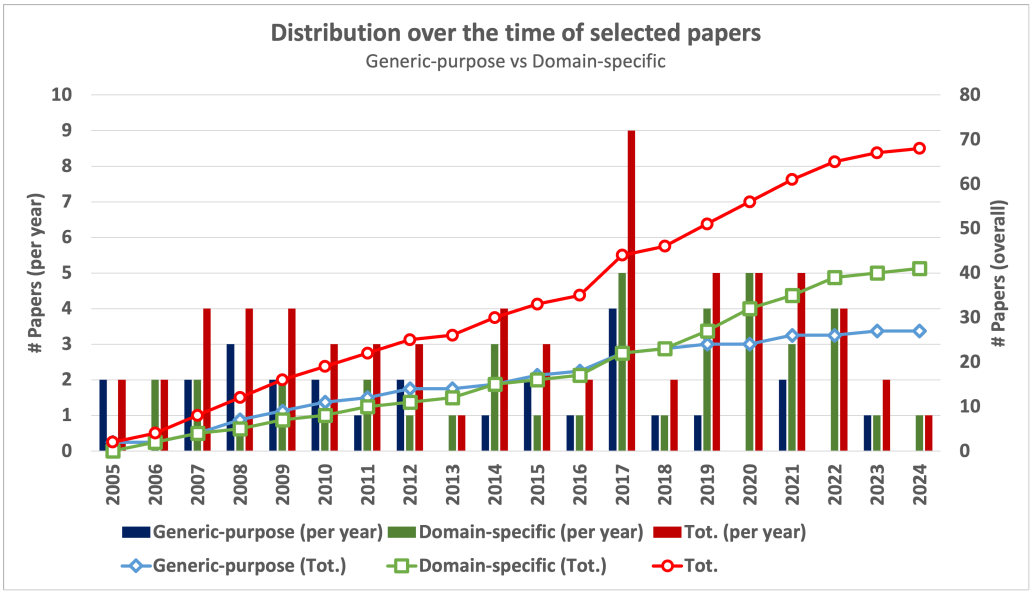
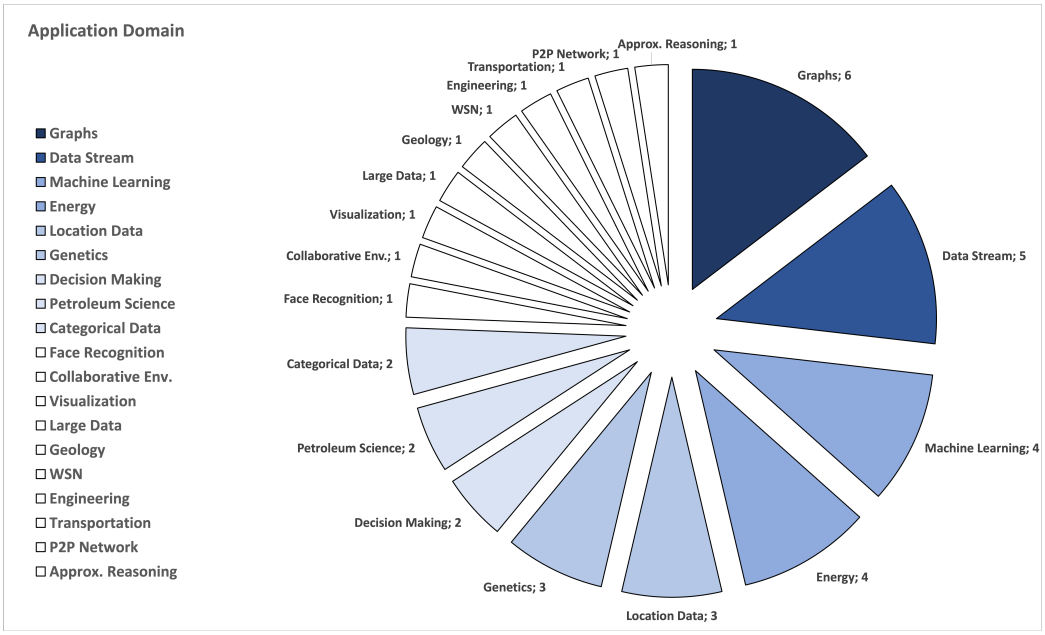


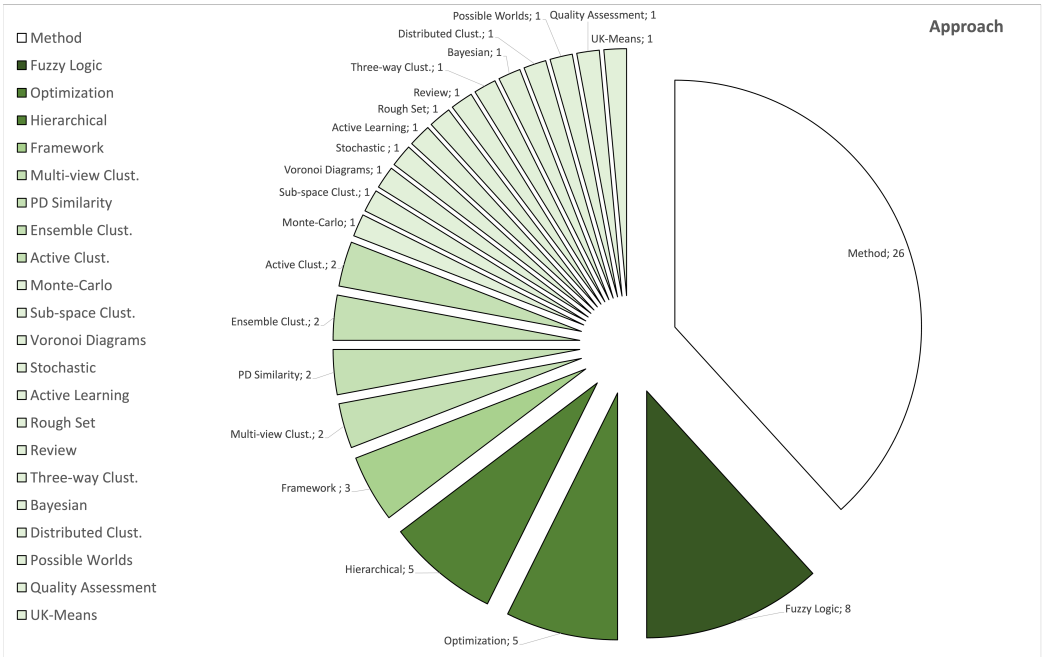
Figure 4. The distribution over the time of the selected contributions.

The analysis conducted in this study, based on soft classification, provides us with an overview of the application domain (Figure 5a). Looking at the 41 domain-specific papers, as expected, generic application fields, such as *graphs*, *data stream*, and *machine learning*, are quantitatively more relevant, both with large domains (e.g., *energy*, *genetics*, and *location data*). At a more fine-grained level, the review has identified a diverse spectrum reflecting a generic need for clustering in the presence of uncertainty.



(**a**) Application domain.

Figure 5. *Cont.*



(**b**) Approaches characterizing the selected solutions.

Figure 5. Analysis overview.

A more technical perspective is summarised in Figure 5b. A consistent amount (38%) of the considered papers propose a mixed-method approach, which is generically referred to as *method* in the adopted analysis framework. While potentially valuable in a domain specific context from an application perspective, in general terms these works are less prone to generalisation or actual innovation in the field. *Fuzzy Logic*, *Optimization*, and *Hierarchical Clustering* are the most popular approaches. They intrinsically constitute the backbone of the identified body of knowledge by providing reference solutions in the field. In addition, to note, a focus has been placed on *analysis frameworks*, on *Multi-view Clustering*, *Probability Distribution Similarity*, *Ensemble* and *Active Clustering*. This last set of works may be understood as a further consolidation by providing a variety of techniques and methods to adapt to the different applications.

5.2. Major Gaps and Challenges

From a critical perspective, the analysis conducted has allowed fpr the identification of a number of gaps other than those originally reported in the different contributions that are summarised in Table 3.

Table 3. Main gaps.

	Gap
G1	Lack of freely available implementations to provide ready-to-use computational resources.
G2	The relationship between generic-purpose and domain-specific solutions os not always clear, namely, ad-hoc solutions do not always place an emphasis on their characteristics and peculiarities.
G3	A fine-grained application-specific approach that does not facilitate re-use in a different context.
G4	There is a tangible lack of abstraction in domain-specific approaches, which often focus on a specific problem without formally defining it. This does not allow for reasoning in terms of the classes of problems.
G5	Despite the existence of a well-identified class of methods and techniques, the solutions proposed in the different works are not always discussed in context looking at the existing body of knowledge, namely the already existing approaches.
G6	The propagation of uncertainty across the different steps in the current technological context characterised by data and computational intensive solutions is not fully analysed and critically discussed.

The review has reiterated the practical relevance of clustering in presence of uncertainty. In such a context, ready-to-use resources in the computational world are crucial and a determinant to consolidate and properly transfer innovation into practice (*G1*). Classic algorithms are implemented as part of more generic computational packages, while possible alternative approaches discussed in the literature are not converging towards more specific computational libraries.

The cross-domain focus has highlighted and put emphasis on applications to solve real-world problems. The relationship between generic-purpose and domain-specific solutions is not always clear (*G2*). The fine-grained application-specific approach makes re-use complex and costly (*G3*). This is because of a lack of abstraction in the formulation of domain-specific problems (*G4*) with a consequent difficulty in generalizing solutions or re-using existing ones in a different context. More in general, despite a well-identified research field, solutions are not always discussed in context, looking at the existing body of knowledge (*G5*).

Last but not least, the propagation of the uncertainty in the current technological context characterised by data and computational intensive solutions is not fully analysed and critically discussed in the literature (*G6*).

5.3. Consolidation and Operationalization Through Open Software

Looking more holistically at the analysis conducted, from a more conceptual perspective that is consistent and fundamentally in line with another call for Open Science [105]. In this specific case, the focus is on computational resources [106], so mostly on Open Software.

The potentially critical role of Open Science, open data and Open Software in the research landscape has been extensively discussed in the literature in general terms, as well as within specific domains (e.g., energy research [107]). It applies also to modern computational trends, such as machine learning, which enables an intersection of computer science and statistics to build systems able to dynamically evolve through experience [108]. One of the key factors underlying the recent advances in the field, as well as their application to solve real world problems and develop advanced systems, is the availability of open-source computational resources. This is perfectly aligned, in concept and practice, with Open Science principles.

These considerations may definitely be generic. In this specific case, looking at the research conducted, there is a tangible feeling that an approach more in line with Open Science would probably allow for a faster and more effective consolidation of the existing body of knowledge in the field, a more effective re-use and application of existing solutions, as well as enhanced support for further evolution. Such a consolidation should facilitate and enhance the operationalization of the different solutions by turning research outcomes into computational resources effectively available for the community.

5.4. Clustering in a Data- and Computationally Intensive Society: Uncertainty Propagation

The intent and extent of clustering within the modern computational context may vary significantly from case to case. While clustering techniques are extensively used in relatively simple contexts, in general terms, it is reasonable to assume that in advanced computational and data-intensive applications clustering is actually part of a multi-stage process.

Moreover, because of its inherent characteristic, it is often adopted in early stages within complex solutions with a realistic assumption that the clustering output may be the input of another contextual step. Therefore, there is an intrinsic risk to propagate uncertainty issues at later stages, with a concrete impact that depends on the sensitivity of the process or application. It is, for instance, the case with data pipelines, whose complexity is constantly increasing, and decision-making processes, which often involve abstractions

progressively built from lower-level data. It underscores the need for a proper uncertainty modelling and management in clustering.

Classic examples that may present issues in terms of uncertainty propagation include, among others, *hybrid approaches to data analysis*, where analysis is conducted applying multiple techniques (e.g., [109]), *exploratory research* (e.g., [110]), *data classification* [111] and *multi-stage machine learning* (e.g., [112]).

6. Conclusions

Given the popularity of clustering techniques within the modern computational world and the intrinsic need to deal with uncertainty in the different application domains, this concise literature review has provided a cross-domain analysis of the most recent solutions in the field.

Such an analysis has underscored the relevance of the topic and the consequently related research activity. A trend towards domain-specific solutions over generic-purpose approaches seems to be dominant and has become more consistent in the last few years. On one hand, this trend enables a more specific set of solutions within specific communities; on the other hand, the resulting distributed approach is not always well integrated in the mainstream and may generate a further fragmentation of the body of knowledge (understood as accepted knowledge and skills required in a specific field or industry), mostly because of some lack of abstraction in the definition of specific problems. Indeed, looking at the specific field of clustering in the presence of uncertainty, such knowledge is fragmented, as it is not always possible to understand how the different solutions are related to each other and how they perform in a given application context.

While these gaps are largely understandable within the research community, addressing the lack of implementations to provide ready-to-use resources is critical overall, looking at a more and more computational and data intensive world.

More holistically, this research has critically addressed the need for approaches more aligned with an open philosophy and the provided considerations in context, looking at the current computational trends that suggest a high risk of uncertainty propagation within complex solutions.

Future analysis steps could be conducted according to a more abstracted and problem-centric framework. Such an approach should be understood as a natural extension of the cross-domain perspective object of this work, which should partially overcome some major limitations by enabling a more qualitative analysis.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: This is a literature review. The considered papers are reported in the paper.

Acknowledgments: The author would like to acknowledge the extensive and constructive feedback provided by the three anonymous reviewers which resulted in a concrete improvement of the paper.

Conflicts of Interest: The authors declare no conflicts of interest.

References

1. Cormode, G.; McGregor, A. Approximation algorithms for clustering uncertain data. In Proceedings of the 27th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Seattle, WA, USA, 18–23 June 2008; pp. 191–200.
2. Weng, C.H.; Chen, Y.L. Mining fuzzy association rules from uncertain data. *Knowl. Inf. Syst.* **2010**, *23*, 129–152. [CrossRef]

3. D'Urso, P. Informational Paradigm, management of uncertainty and theoretical formalisms in the clustering framework: A review. *Inf. Sci.* **2017**, *400*, 30–62. [CrossRef]
4. Kuczenski, B. False confidence: Are we ignoring significant sources of uncertainty? *Int. J. Life Cycle Assess.* **2019**, *24*, 1760–1764. [CrossRef]
5. Griffin, S.C.; Claxton, K.P.; Palmer, S.J.; Sculpher, M.J. Dangerous omissions: The consequences of ignoring decision uncertainty. *Health Econ.* **2011**, *20*, 212–224. [CrossRef] [PubMed]
6. Brodlie, K.; Allendes Osorio, R.; Lopes, A. A review of uncertainty in data visualization. In *Expanding the Frontiers of Visual Analytics and Visualization*; Springer: Berlin/Heidelberg, Germany, 2012; pp. 81–109.
7. Nuijten, M.; Mittendorf, T.; Persson, U. Practical issues in handling data input and uncertainty in a budget impact analysis. *Eur. J. Health Econ.* **2011**, *12*, 231–241. [CrossRef]
8. Karimi, J.; Somers, T.M.; Gupta, Y.P. Impact of environmental uncertainty and task characteristics on user satisfaction with data. *Inf. Syst. Res.* **2004**, *15*, 175–193. [CrossRef]
9. McMillan, H.K.; Westerberg, I.K.; Krueger, T. Hydrological data uncertainty and its implications. *Wiley Interdiscip. Rev. Water* **2018**, *5*, e1319. [CrossRef]
10. Hariri, R.H.; Fredericks, E.M.; Bowers, K.M. Uncertainty in big data analytics: Survey, opportunities, and challenges. *J. Big Data* **2019**, *6*, 44. [CrossRef]
11. Wang, X.; He, Y. Learning from uncertainty for big data: Future analytical challenges and strategies. *IEEE Syst. Man Cybern. Mag.* **2016**, *2*, 26–31. [CrossRef]
12. Kamal, A.; Dhakal, P.; Javaid, A.Y.; Devabhaktuni, V.K.; Kaur, D.; Zaientz, J.; Marinier, R. Recent advances and challenges in uncertainty visualization: A survey. *J. Vis.* **2021**, *24*, 861–890. [CrossRef]
13. Abdar, M.; Pourpanah, F.; Hussain, S.; Rezazadegan, D.; Liu, L.; Ghavamzadeh, M.; Fieguth, P.; Cao, X.; Khosravi, A.; Acharya, U.R.; et al. A review of uncertainty quantification in deep learning: Techniques, applications and challenges. *Inf. Fusion* **2021**, *76*, 243–297. [CrossRef]
14. Xu, R.; Wunsch, D. Survey of clustering algorithms. *IEEE Trans. Neural Netw.* **2005**, *16*, 645–678. [CrossRef] [PubMed]
15. Wierzchoń, S.T.; Kłopotek, M.A. *Modern Algorithms of Cluster Analysis*; Springer: Berlin/Heidelberg, Germany, 2018; Volume 34.
16. Sinaga, K.P.; Yang, M.S. Unsupervised K-means clustering algorithm. *IEEE Access* **2020**, *8*, 80716–80727. [CrossRef]
17. Caron, M.; Bojanowski, P.; Joulin, A.; Douze, M. Deep clustering for unsupervised learning of visual features. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 132–149.
18. Castellanos, A.; Cigarrán, J.; García-Serrano, A. Formal concept analysis for topic detection: A clustering quality experimental analysis. *Inf. Syst.* **2017**, *66*, 24–42. [CrossRef]
19. Lee, C.S.; Kao, Y.F.; Kuo, Y.H.; Wang, M.H. Automated ontology construction for unstructured text documents. *Data Knowl. Eng.* **2007**, *60*, 547–566. [CrossRef]
20. Pileggi, S.F. Ontological Modelling and Social Networks: From Expert Validation to Consolidated Domains. In *Lectures Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2023; pp. 672–687.
21. Tew, C.; Giraud-Carrier, C.; Tanner, K.; Burton, S. Behavior-based clustering and analysis of interestingness measures for association rule mining. *Data Min. Knowl. Discov.* **2014**, *28*, 1004–1045. [CrossRef]
22. Jiang, B.; Pei, J.; Tao, Y.; Lin, X. Clustering uncertain data based on probability distribution similarity. *IEEE Trans. Knowl. Data Eng.* **2011**, *25*, 751–763. [CrossRef]
23. Hüllermeier, E. Uncertainty in clustering and classification. In Proceedings of the Scalable Uncertainty Management: 4th International Conference, SUM 2010, Toulouse, France, 27–29 September 2010; Springer: Berlin/Heidelberg, Germany, 2010; pp. 16–19.
24. de Kok, J.W.T.M.; van Rosmalen, F.; Koeze, J.; Keus, F.; van Kuijk, S.M.J.; Forte, J.C.; Schnabel, R.M.; Driessen, R.G.H.; van Herpt, T.T.W.; Sels, J.-W.E.M.; et al. Deep embedded clustering generalisability and adaptation for integrating mixed datatypes: Two critical care cohorts. *Sci. Rep.* **2024**, *14*, 1045. [CrossRef]
25. Scarselli, F.; Gori, M.; Chung Tsoi, A.; Hagenbuchner, M.; Monfardini, G. The Graph Neural Network Model. *IEEE Trans. Neural Netw. Learn. Syst.* **2009**, *20*, 61–80. [CrossRef]
26. Aggarwal, C.C.; Philip, S.Y. A survey of uncertain data algorithms and applications. *IEEE Trans. Knowl. Data Eng.* **2008**, *21*, 609–623. [CrossRef]
27. Pileggi, S.F. A Cross-Domain Perspective to Clustering with Uncertainty. In *International Conference on Computational Science*; Springer: Cham, Switzerland, 2024; pp. 295–308.
28. Cheng, Y.; Fu, K.S. Conceptual clustering in knowledge organization. *IEEE Trans. Pattern Anal. Mach. Intell.* **1985**, *5*, 592–598. [CrossRef]
29. Lee, R.C. Clustering analysis and its applications. In *Advances in Information Systems Science: Volume 8*; Springer: Berlin/Heidelberg, Germany, 1981; pp. 169–292.
30. Thorndike, R.L. Who belongs in the family? *Psychometrika* **1953**, *18*, 267–276. [CrossRef]

31. Saxena, A.; Prasad, M.; Gupta, A.; Bharill, N.; Patel, O.P.; Tiwari, A.; Er, M.J.; Ding, W.; Lin, C.T. A review of clustering techniques and developments. *Neurocomputing* **2017**, *267*, 664–681. [CrossRef]
32. Yang, M.S. A survey of fuzzy clustering. *Math. Comput. Model.* **1993**, *18*, 1–16. [CrossRef]
33. Zhang, C.; Butepage, J.; Kjellstrom, H.; Mandt, S. Advances in Variational Inference. *IEEE Trans. Pattern Anal. Mach. Intell.* **2019**, *41*, 2008–2026. [CrossRef]
34. Liu, Y.; Liu, Z.; Li, S.; Guo, Y.; Liu, Q.; Wang, G. Cloud-Cluster: An uncertainty clustering algorithm based on cloud model. *Knowl.-Based Syst.* **2023**, *263*, 110261. [CrossRef]
35. Sharma, K.K.; Seal, A. Outlier-robust multi-view clustering for uncertain data. *Knowl.-Based Syst.* **2021**, *211*, 106567. [CrossRef]
36. Sharma, K.K.; Seal, A. Multi-view spectral clustering for uncertain objects. *Inf. Sci.* **2021**, *547*, 723–745. [CrossRef]
37. Sharma, K.K.; Seal, A. Modeling uncertain data using Monte Carlo integration method for clustering. *Expert Syst. Appl.* **2019**, *137*, 100–116. [CrossRef]
38. Dalton, L.A.; Benalcázar, M.E.; Dougherty, E.R. Optimal clustering under uncertainty. *PLoS ONE* **2018**, *13*, e0204627. [CrossRef]
39. Huang, D.; Wang, C.D.; Lai, J.H. Locally weighted ensemble clustering. *IEEE Trans. Cybern.* **2017**, *48*, 1460–1473. [CrossRef] [PubMed]
40. Liu, H.; Zhang, X.; Zhang, X.; Cui, Y. Self-adapted mixture distance measure for clustering uncertain data. *Knowl.-Based Syst.* **2017**, *126*, 33–47. [CrossRef]
41. Zhang, X.; Liu, H.; Zhang, X. Novel density-based and hierarchical density-based clustering algorithms for uncertain data. *Neural Netw.* **2017**, *93*, 240–255. [CrossRef]
42. Gullo, F.; Ponti, G.; Tagarelli, A.; Greco, S. An information-theoretic approach to hierarchical clustering of uncertain data. *Inf. Sci.* **2017**, *402*, 199–215. [CrossRef]
43. Xiong, C.; Johnson, D.M.; Corso, J.J. Active clustering with model-based uncertainty reduction. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *39*, 5–17. [CrossRef]
44. Huang, D.; Lai, J.H.; Wang, C.D. Robust ensemble clustering using probability trajectories. *IEEE Trans. Knowl. Data Eng.* **2015**, *28*, 1312–1326. [CrossRef]
45. Xu, L.; Hu, Q.; Hung, E.; Chen, B.; Tan, X.; Liao, C. Large margin clustering on uncertain data by considering probability distribution similarity. *Neurocomputing* **2015**, *158*, 81–89. [CrossRef]
46. Züfle, A.; Emrich, T.; Schmid, K.A.; Mamoulis, N.; Zimek, A.; Renz, M. Representative clustering of uncertain data. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Barcelona, Spain, 25–29 August 2014; pp. 243–252.
47. Wauthier, F.L.; Jojic, N.; Jordan, M.I. Active spectral clustering via iterative uncertainty reduction. In Proceedings of the 18th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Beijing, China, 12–16 August 2012; pp. 1339–1347.
48. Gullo, F.; Ponti, G.; Tagarelli, A. Minimizing the variance of cluster mixture models for clustering uncertain objects. *Stat. Anal. Data Mining ASA Data Sci. J.* **2013**, *6*, 116–135. [CrossRef]
49. Kao, B.; Lee, S.D.; Lee, F.K.; Cheung, D.W.; Ho, W.S. Clustering uncertain data using voronoi diagrams and r-tree index. *IEEE Trans. Knowl. Data Eng.* **2010**, *22*, 1219–1233.
50. Günnemann, S.; Kremer, H.; Seidl, T. Subspace clustering for uncertain data. In Proceedings of the 2010 SIAM International Conference on Data Mining, Columbus, OH, USA, 29 April–1 May 2010; pp. 385–396.
51. Guha, S.; Munagala, K. Exceeding expectations and clustering uncertain data. In Proceedings of the 28th ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, Seattle, WA, USA, 18–23 June 2009; pp. 269–278.
52. Volk, P.B.; Rosenthal, F.; Hahmann, M.; Habich, D.; Lehner, W. Clustering uncertain data with possible worlds. In Proceedings of the 2009 IEEE 25th International Conference on Data Engineering, Shanghai, China, 29 March–2 April 2009; pp. 1625–1632.
53. Gullo, F.; Ponti, G.; Tagarelli, A. Clustering uncertain data via k-medoids. In *Lecture Notes on Computer Science*; Springer: Berlin/Heidelberg, Germany, 2008; pp. 229–242.
54. Gullo, F.; Ponti, G.; Tagarelli, A.; Greco, S. A hierarchical algorithm for clustering uncertain data via an information-theoretic approach. In Proceedings of the 2008 28th IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 821–826.
55. Kao, B.; Lee, S.D.; Cheung, D.W.; Ho, W.S.; Chan, K. Clustering uncertain data using voronoi diagrams. In Proceedings of the 2008 28th IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; pp. 333–342.
56. Rhee, F.C.H. Uncertain fuzzy clustering: Insights and recommendations. *IEEE Comput. Intell. Mag.* **2007**, *1*, 44–56.
57. Hwang, C.; Rhee, F.C.H. Uncertain fuzzy clustering: Interval type-2 fuzzy approach to *c*-means. *IEEE Trans. Fuzzy Syst.* **2007**, *15*, 107–120. [CrossRef]
58. Kriegel, H.P.; Pfeifle, M. Density-based clustering of uncertain data. In Proceedings of the 11th ACM SIGKDD International Conference on Knowledge Discovery in Data Mining, Chicago, IL, USA, 21–24 August 2005; pp. 672–677.
59. Kriegel, H.P.; Pfeifle, M. Hierarchical density-based clustering of uncertain data. In Proceedings of the 5th IEEE International Conference on Data Mining (ICDM'05), Houston, TX, USA, 27–30 November 2005; p. 4.

60. Bhavsar, S.; Pitchumani, R.; Maack, J.; Satkauskas, I.; Reynolds, M.; Jones, W. Stochastic economic dispatch of wind power under uncertainty using clustering-based extreme scenarios. *Electr. Power Syst. Res.* **2024**, *229*, 110158. [CrossRef]
61. Rendon, N.; Giraldo, J.H.; Bouwmans, T.; Rodríguez-Buritica, S.; Ramirez, E.; Isaza, C. Uncertainty clustering internal validity assessment using Fréchet distance for unsupervised learning. *Eng. Appl. Artif. Intell.* **2023**, *124*, 106635. [CrossRef]
62. He, Y.; Yang, J.P.; Li, Y.F. A three-stage automated modal identification framework for bridge parameters based on frequency uncertainty and density clustering. *Eng. Struct.* **2022**, *255*, 113891. [CrossRef]
63. Hussain, S.F.; Butt, I.A.; Hanif, M.; Anwar, S. Clustering uncertain graphs using ant colony optimization (ACO). *Neural Comput. Appl.* **2022**, *34*, 11721–11738. [CrossRef]
64. Wang, P.; Ding, C.; Tan, W.; Gong, M.; Jia, K.; Tao, D. Uncertainty-aware clustering for unsupervised domain adaptive object re-identification. *IEEE Trans. Multimed.* **2022**, *25*, 2624–2635. [CrossRef]
65. Hewitt, M.; Ortmann, J.; Rei, W. Decision-based scenario clustering for decision-making under uncertainty. *Ann. Oper. Res.* **2022**, *315*, 747–771. [CrossRef]
66. Prabhu, V.; Chandrasekaran, A.; Saenko, K.; Hoffman, J. Active domain adaptation via clustering uncertainty-weighted embeddings. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 8505–8514.
67. Debnath, B.; Coviello, G.; Yang, Y.; Chakradhar, S. UAC: An uncertainty-aware face clustering algorithm. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3487–3495.
68. Haddadpour, H.; Niri, M.E. Uncertainty assessment in reservoir performance prediction using a two-stage clustering approach: Proof of concept and field application. *J. Pet. Sci. Eng.* **2021**, *204*, 108765. [CrossRef]
69. Shi, W.; Chen, W.N.; Gu, T.; Jin, H.; Zhang, J. Handling uncertainty in financial decision making: A clustering estimation of distribution algorithm with simplified simulation. *IEEE Trans. Emerg. Top. Comput. Intell.* **2020**, *5*, 42–56. [CrossRef]
70. Li, Y.; Chung, S.H. Ride-sharing under travel time uncertainty: Robust optimization and clustering approaches. *Comput. Ind. Eng.* **2020**, *149*, 106601. [CrossRef]
71. Huang, J.; Gong, S.; Zhu, X. Deep semantic clustering by partition confidence maximisation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8849–8858.
72. Naouali, S.; Salem, S.B.; Chtourou, Z. Uncertainty mode selection in categorical clustering using the rough set theory. *Expert Syst. Appl.* **2020**, *158*, 113555. [CrossRef]
73. Charwand, M.; Gitizadeh, M.; Siano, P.; Chicco, G.; Moshavash, Z. Clustering of electrical load patterns and time periods using uncertainty-based multi-level amplitude thresholding. *Int. J. Electr. Power Energy Syst.* **2020**, *117*, 105624. [CrossRef]
74. Kang, B.; Kim, S.; Jung, H.; Choe, J.; Lee, K. Efficient assessment of reservoir uncertainty using distance-based clustering: A review. *Energies* **2019**, *12*, 1859. [CrossRef]
75. Shukla, A.K.; Muhuri, P.K. Big-data clustering with interval type-2 fuzzy uncertainty modeling in gene expression datasets. *Eng. Appl. Artif. Intell.* **2019**, *77*, 268–282. [CrossRef]
76. Tabesh, M.; Askari-Nasab, H. Clustering mining blocks in presence of geological uncertainty. *Min. Technol.* **2019**, *128*, 162–176. [CrossRef]
77. Han, K.; Gui, F.; Xiao, X.; Tang, J.; He, Y.; Cao, Z.; Huang, H. Efficient and effective algorithms for clustering uncertain graphs. *Proc. VLDB Endow.* **2019**, *12*, 667–680. [CrossRef]
78. Afridi, M.K.; Azam, N.; Yao, J.; Alanazi, E. A three-way clustering approach for handling missing data using GTRS. *Int. J. Approx. Reason.* **2018**, *98*, 11–24. [CrossRef]
79. Ceccarello, M.; Fantozzi, C.; Pietracaprina, A.; Pucci, G.; Vandin, F. Clustering uncertain graphs. *Proc. VLDB Endow.* **2017**, *11*, 472–484. [CrossRef]
80. Yao, C.; Chen, M.; Hong, Y.Y. Novel adaptive multi-clustering algorithm-based optimal ESS sizing in ship power system considering uncertainty. *IEEE Trans. Power Syst.* **2017**, *33*, 307–316. [CrossRef]
81. Chang, Y.; Chen, J.; Cho, M.H.; Castaldi, P.J.; Silverman, E.K.; Dy, J.G. Multiple clustering views from multiple uncertain experts. In Proceedings of the International Conference on Machine Learning, PMLR, Sydney, Australia, 6–11 August 2017; pp. 674–683.
82. Zhou, J.; Chen, L.; Chen, C.P.; Wang, Y.; Li, H.X. Uncertain data clustering in distributed peer-to-peer networks. *IEEE Trans. Neural Netw. Learn. Syst.* **2017**, *29*, 2392–2406. [CrossRef]
83. Halim, Z.; Waqas, M.; Baig, A.R.; Rashid, A. Efficient clustering of large uncertain graphs using neighborhood information. *Int. J. Approx. Reason.* **2017**, *90*, 274–291. [CrossRef]
84. Shukla, A.; Singh, S. Clustering based unit commitment with wind power uncertainty. *Energy Convers. Manag.* **2016**, *111*, 89–102. [CrossRef]
85. Schubert, E.; Koos, A.; Emrich, T.; Züfle, A.; Schmid, K.A.; Zimek, A. A framework for clustering uncertain data. *Proc. VLDB Endow.* **2015**, *8*, 1976–1979. [CrossRef]
86. Jin, C.; Yu, J.X.; Zhou, A.; Cao, F. Efficient clustering of uncertain data streams. *Knowl. Inf. Syst.* **2014**, *40*, 509–539. [CrossRef]

87. Luo, Q.; Peng, Y.; Peng, X.; Saddik, A.E. Uncertain data clustering-based distance estimation in wireless sensor networks. *Sensors* **2014**, *14*, 6584–6605. [CrossRef]
88. Chen, Y.; Lim, S.H.; Xu, H. Weighted graph clustering with non-uniform uncertainties. In Proceedings of the International Conference on Machine Learning. PMLR, Beijing, China, 21–26 June 2014; pp. 1566–1574.
89. Ghosh, S.; Mitra, S. Clustering large data with uncertainty. *Appl. Soft Comput.* **2013**, *13*, 1639–1645. [CrossRef]
90. Liu, L.; Jin, R.; Aggarwal, C.; Shen, Y. Reliable clustering on uncertain graphs. In Proceedings of the 2012 IEEE 12th International Conference on Data Mining, Brussels, Belgium, 10 December 2012; pp. 459–468.
91. Pelekis, N.; Kopanakis, I.; Kotsifakos, E.E.; Frentzos, E.; Theodoridis, Y. Clustering uncertain trajectories. *Knowl. Inf. Syst.* **2011**, *28*, 117–147. [CrossRef]
92. Meesuksabai, W.; Kangkachit, T.; Waiyamai, K. Hue-stream: Evolution-based clustering technique for heterogeneous data streams with uncertainty. In Proceedings of the Advanced Data Mining and Applications: 7th International Conference, ADMA 2011, Beijing, China, 17–19 December 2011; Springer: Berlin/Heidelberg, Germany, 2011; pp. 27–40.
93. Huang, G.Y.; Liang, D.P.; Hu, C.Z.; Ren, J.D. An algorithm for clustering heterogeneous data streams with uncertainty. In Proceedings of the 2010 International Conference on Machine Learning and Cybernetics, Qingdao, China, 11–14 July 2010; Volume 4, pp. 2059–2064.
94. Aggarwal, C.C. On high dimensional projected clustering of uncertain data streams. In Proceedings of the 2009 IEEE 25th International Conference on Data Engineering, Shanghai, China, 29 March–2 April 2009; pp. 1152–1154.
95. Pelekis, N.; Kopanakis, I.; Kotsifakos, E.; Frentzos, E.; Theodoridis, Y. Clustering trajectories of moving objects in an uncertain world. In Proceedings of the 2009 9th IEEE international Conference on Data Mining, Miami Beach, FL, USA, 6–9 December 2009; pp. 417–427.
96. Aggarwal, C.C.; Philip, S.Y. A framework for clustering uncertain data streams. In Proceedings of the 2008 IEEE 24th International Conference on Data Engineering, Cancun, Mexico, 7–12 April 2008; pp. 150–159.
97. Xia, Y.; Xi, B. Conceptual clustering categorical data with uncertainty. In Proceedings of the 19th IEEE International Conference on Tools with Artificial Intelligence (ICTAI 2007), Patras, Greece, 29–31 October 2007; Volume 1, pp. 329–336.
98. Liu, X.; Lin, K.K.; Andersen, B.; Rattray, M. Including probe-level uncertainty in model-based gene expression clustering. *BMC Bioinform.* **2007**, *8*, 98. [CrossRef]
99. Suzuki, R.; Shimodaira, H. Pvclust: An R package for assessing the uncertainty in hierarchical clustering. *Bioinformatics* **2006**, *22*, 1540–1542. [CrossRef]
100. Chau, M.; Cheng, R.; Kao, B.; Ng, J. Uncertain data mining: An example in clustering location data. In Proceedings of the Advances in Knowledge Discovery and Data Mining: 10th Pacific-Asia Conference, PAKDD 2006, Singapore, 9–12 April 2006; pp. 199–204.
101. Ran, X.; Xi, Y.; Lu, Y.; Wang, X.; Lu, Z. Comprehensive survey on hierarchical clustering algorithms and the recent developments. *Artif. Intell. Rev.* **2023**, *56*, 8219–8264. [CrossRef]
102. Boongoen, T.; Iam-On, N. Cluster ensembles: A survey of approaches with recent extensions and applications. *Comput. Sci. Rev.* **2018**, *28*, 1–25. [CrossRef]
103. Fu, L.; Lin, P.; Vasilakos, A.V.; Wang, S. An overview of recent multi-view clustering. *Neurocomputing* **2020**, *402*, 148–161. [CrossRef]
104. Ibrahim, D. An Overview of Soft Computing. *Procedia Comput. Sci.* **2016**, *102*, 34–38. [CrossRef]
105. Vicente-Saez, R.; Martinez-Fuentes, C. Open Science now: A systematic literature review for an integrated definition. *J. Bus. Res.* **2018**, *88*, 428–436. [CrossRef]
106. Bonaccorsi, A.; Rossi, C. Why open source software can succeed. *Res. Policy* **2003**, *32*, 1243–1258. [CrossRef]
107. Pfenninger, S.; DeCarolis, J.; Hirth, L.; Quoilin, S.; Staffell, I. The importance of open data and software: Is energy research lagging behind? *Energy Policy* **2017**, *101*, 211–215. [CrossRef]
108. Jordan, M.I.; Mitchell, T.M. Machine learning: Trends, perspectives, and prospects. *Science* **2015**, *349*, 255–260. [CrossRef]
109. Fan, C.Y.; Fan, P.S.; Chan, T.Y.; Chang, S.H. Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Syst. Appl.* **2012**, *39*, 8844–8851. [CrossRef]
110. Pileggi, S.F. A hybrid approach to analysing large scale surveys: Individual values, opinions and perceptions. *SN Soc. Sci.* **2024**, *4*, 144. [CrossRef]
111. Oyewole, G.J.; Thopil, G.A. Data clustering: Application and trends. *Artif. Intell. Rev.* **2023**, *56*, 6439–6475. [CrossRef]
112. Mardani, A.; Liao, H.; Nilashi, M.; Alrasheedi, M.; Cavallaro, F. A multi-stage method to predict carbon dioxide emissions using dimensionality reduction, clustering, and machine learning techniques. *J. Clean. Prod.* **2020**, *275*, 122942. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.