



## RESEARCH ARTICLE

# Teaching Case Study in Discrete Mathematics: Application of Clustering Algorithms in Brucella Traceability Research

Jiqing Gu<sup>1</sup>, Jing Hu<sup>1</sup>, Ju Huang<sup>1</sup> and Hui Yang<sup>1,2,\*</sup>

<sup>1</sup>School of Computer Science, Chengdu University of Information Technology, China

<sup>2</sup>Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, China

**Abstract:** In this study, we combined discrete mathematics with computational biology to identify Brucella species using clustering algorithms. By analyzing the Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) spectra of 44 Brucella isolates, which included 21 Brucella melitensis, 12 Brucella suis, and 11 Brucella abortus, we utilized a feature selection strategy to pinpoint 22 key characteristics critical for species differentiation. We then developed a spectral clustering-based model for Brucella traceability. This model offers a rapid and cost-effective alternative to traditional, labor-intensive identification methods, significantly improving the efficiency and accuracy of Brucella strain identification. Additionally, it aids in monitoring disease transmission trends, identifying outbreak sources, and formulating effective control strategies to mitigate risks. Our findings demonstrate the practical application of discrete mathematics in computational biology, contributing significantly to both scientific research and educational methodologies. This approach illustrates how mathematical concepts can be effectively applied to solve real-world biological problems, providing valuable insights for future interdisciplinary studies and innovative solutions.

**Keywords:** Brucella, mass spectrometry, clustering, discrete mathematics

## 1. Introduction

Discrete mathematics plays a crucial role in the education of mathematical modeling in computational biology. It encompasses various areas such as combinatorics, number theory, and graph theory, providing an ideal framework for exploring and solving practical problems [1]. Moreover, combining discrete mathematics with mathematical modeling strengthens its practical application in modeling activities, allowing students to gain a deeper understanding of this field [2]. This integration brings new opportunities for teaching, especially in the context of mathematical modeling, where the knowledge of discrete mathematics is not only viewed as a theoretical tool but also plays a significant role in practical applications [3, 4]. Thus, the application of discrete mathematics in mathematical modeling education, especially in the field of computational biology, not only enhances the practicality and realism of the teaching content but also promotes the development of students' critical skills, presenting important opportunities and challenges for modern mathematics education.

Cluster analysis is crucial in studying Brucella species, especially in epidemiology. Traditional microbial isolation and typing procedures for Brucella strains are expensive and

time-consuming, requiring biohazard control facilities, which makes the development of a rapid and effective identification method a pressing need [5]. Cluster analysis, as a computational approach, offers the ability to rapidly classify and type a large number of Brucella samples, thereby providing important information for determining the sources and transmission pathways of the pathogen. Through cluster analysis, the epidemiological characteristics of Brucella strains can be assessed, aiding in determining the relationships and differences between different strains [6]. This method enables researchers to more effectively monitor disease transmission trends, understand the mechanisms of transmission, and thus formulate more targeted control and prevention strategies. Overall, the application of computational clustering methods in Brucella research not only improves the efficiency and accuracy of pathogen identification but also has significant value for public health decision-making and epidemiological investigations. The insights from Brucella traceability can be applied to other infectious diseases by using similar tracking methods, such as genomic sequencing and advanced clustering algorithms, to trace the spread of pathogens, identify transmission routes, and predict outbreaks. This approach can enhance surveillance, improve early detection, and support targeted control measures for better disease management.

This study analyzed the Matrix-Assisted Laser Desorption/Ionization Time of Flight Mass Spectrometry (MALDI-TOF MS) spectra of 44 Brucella isolates (including 21 B. melitensis, 12 B. suis, and 11 B. abortus) using clustering algorithms, identified 29

\*Corresponding author: Hui Yang, School of Computer Science, Chengdu University of Information Technology and Yangtze Delta Region Institute (Quzhou), University of Electronic Science and Technology of China, China. Email: [huiyang@csj.uestc.edu.cn](mailto:huiyang@csj.uestc.edu.cn)

optimal characteristics, and established a clustering algorithm that can be used for the traceability of Brucella.

## 2. Theoretical Background

In the field of data clustering, the application of discrete mathematics is extremely widespread, providing key theoretical support for understanding and implementing various clustering algorithms. Core concepts such as graph theory, combinatorics, and discrete optimization play a crucial role in different clustering methods.

Graph theory is particularly important in spectral clustering and agglomerative clustering [7, 8]. Spectral clustering uses principles of graph theory, identifying natural groupings of data by analyzing the Laplacian matrix of a graph and its eigenvectors [9]. On the other hand, agglomerative clustering creates a layered, tree-like structure, similar to the tree structures in graph theory, to visually display the relationships between data points [10]. This method builds an agglomerative structure by progressively merging or dividing data points, providing a way to understand the organization of data at different levels. In K-means clustering, combinatorics plays a key role [11]. This clustering method involves determining which data points should be allocated to specific clusters, which is a typical combinatorial problem. Below, we will provide detailed introductions to spectral clustering, K-means clustering, and agglomerative clustering.

## 3. Model Training

### 3.1. Spectral clustering

The spectral clustering algorithm begins with the construction of a similarity matrix  $S$  for a given dataset, where  $S_{ij}$  denotes the similarity between points  $x_i$  and  $x_j$ . It then involves the construction of the graph's degree matrix  $D$  and the adjacency matrix  $A$  (where  $A = S$ ), followed by the computation of the Laplacian matrix  $L = D - A$ . The next step is the eigenvalue decomposition of the Laplacian matrix  $L$ . Finally, the eigenvectors corresponding to the first  $k$  smallest eigenvalues are selected to form a new feature matrix, and clustering is performed by treating each row of this matrix as a point and applying the K-means algorithm. Similarity matrix:

$$S_{ij} = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (1)$$

Degree matrix:

$$D_{ii} = \sum_j A_{ij} \quad (2)$$

Laplacian matrix:

$$L = D - A \quad (3)$$

### 3.2. K-means clustering

The K-means clustering algorithm starts with the random selection of  $k$  data points as initial centroids. Each data point is then assigned to the nearest centroid, forming  $k$  clusters. The centroids of these clusters are recalculated, and the assignment and updating steps are repeated until the centroids stabilize and no longer change significantly.

Centroid update:

$$C_j = \frac{1}{|S_j|} \sum_{x_i \in S_j} x_i \quad (4)$$

where  $S_j$  is the  $j^{\text{th}}$  cluster and  $C_j$  is its centroid.

### 3.3. Agglomerative clustering

The agglomerative clustering algorithm begins by treating each data point as an individual cluster. It then finds and merges the two closest clusters, updating the distances between clusters after each merge. This process is repeated until all data points are included in a single cluster or a predetermined number of clusters is reached.

Cluster distance calculation method (e.g., single linkage):

$$D(S_1, S_2) = \min\{d(x, y) : x \in S_1, y \in S_2\} \quad (5)$$

where  $d(x, y)$  is the distance between points  $x$  and  $y$ .

## 4. Performance Evaluation

In clustering analysis, evaluation metrics are generally divided into external evaluation metrics and internal evaluation metrics, which assess the quality of clustering from different perspectives. External evaluation metrics rely on true labels or external reference information for evaluation and are typically used to assess the similarity between clustering results and true classification labels. Internal evaluation metrics do not depend on true labels but instead evaluate the quality of clustering based on the characteristics of the clustering results themselves, such as cohesion and separation. In this study, the external evaluation metrics include: Adjusted Rand Index (ARI) and Normalized Mutual Information (NMI). The internal evaluation metrics include Silhouette Score, Calinski–Harabasz Index (CHI), and Davies–Bouldin Index (DBI).

### 4.1. Adjusted Rand Index (ARI)

The ARI measures the consistency between two data groupings. Its formula is

$$ARI = \frac{RI - E[RI]}{\max(RI) - E[RI]}$$

where  $RI$  is the Rand Index, representing the similarity between two groupings.  $E[RI]$  is the expected value of the Rand Index for random groupings.  $\max(RI)$  is the maximum possible Rand Index. The ARI ranges from  $-1$  to  $1$ , where  $1$  indicates perfect agreement,  $0$  indicates random grouping, and negative values indicate worse-than-random performance.

### 4.2. Mutual Information (MI)

Mutual Information (MI) measures the amount of information shared between the clustering result and the true labels. Its formula is

$$MI(U, V) = \sum_{i=1}^n \sum_{j=1}^m P(u_i, v_j) \log \frac{P(u_i, v_j)}{P(u_i)P(v_j)}$$

where  $(P(u_i, v_j))$  is the joint probability distribution of true label  $u_i$  and predicted cluster  $v_j$ .

### 4.3. Silhouette coefficient

The silhouette coefficient is a measure of how similar an object is to its own cluster compared to other clusters. The value ranges from  $-1$  to  $1$ , where a value close to  $1$  implies that the clustering is appropriate, while a value close to  $-1$  implies poor clustering.

For a single sample point  $a$ , the silhouette coefficient is calculated as follows:

$$s = \frac{b - a}{\max(a, b)} \quad (6)$$

Here,  $a$  is the mean distance to the other points in the same cluster, and  $b$  is the mean distance to the points in the nearest cluster that the point is not a part of.

### 4.4. Calinski–Harabasz Index (CHI)

Also known as the Variance Ratio Criterion, this index is a ratio of the sum of between-clusters dispersion and of within-cluster dispersion. Higher values of the Calinski–Harabasz score indicate better clustering [12].

Let  $k$  be the number of clusters,  $n$  be the number of samples, and  $B$  and  $W$  be the between-group and within-group dispersion matrices, respectively:

$$CH = \frac{\text{trace}(B)/(k - 1)}{\text{trace}(W)/(n - k)} \quad (7)$$

where  $\text{trace}(B)$  is the trace of the between-group dispersion matrix and  $\text{trace}(W)$  is the trace of the within-group dispersion matrix.

### 4.5. Davies–Bouldin Index (DBI)

This index is defined as the average ratio of within-cluster distances to between-cluster distances. Lower DBI values indicate better clustering [13].

For clusters, the DBI is defined as:

$$DB = \frac{1}{k} \sum_{i=1}^k \max_{j \neq i} \left( \frac{\sigma_i + \sigma_j}{d(c_i, c_j)} \right) \quad (8)$$

Here,  $\sigma_i$  is the average distance of all points in cluster  $i$  to the centroid  $c_i$ , and  $d(c_i, c_j)$  is the distance between centroids of clusters  $i$  and  $j$ .

Each of these metrics offers unique insights into the quality of clustering and is suitable for different scenarios in cluster analysis.

## 5. Case Study Result

### 5.1. The benchmark datasets

MALDI-TOF MS is a soft ionization proteomics platform, widely regarded as the gold standard for microbial fingerprint characterization in clinical laboratories. This technique is known for its repeatability, reliability, and cost-effectiveness, capable of accurately identifying bacteria within minutes at a low cost per reaction. Dematheis et al. used MALDI-TOF MS to analyze three different *Brucella* species identified by the principal component regression method: *Brucella melitensis*, *Brucella abortus*, and *Brucella suis*. The MS spectra were preprocessed and cleaned, creating a peak matrix containing sample IDs and feature peak intensities, with a total of 436 MS peaks identified [14].

We constructed our benchmark dataset with the above data:

$$S = S1 + S2 + S3 \quad (9)$$

It comprises 43 samples in total, each sample containing 436 features.

### 5.2. Feature selection

Feature selection is a key step in data preprocessing, especially when dealing with datasets with a large number of features [15]. Its purpose is to identify the most useful features for modeling in order to improve the performance of the model and reduce computational complexity. By removing irrelevant or redundant features, we can enhance the accuracy and efficiency of the model and simplify its interpretation. The importance of feature selection lies not only in helping to improve the efficiency and performance of the model but also in aiding the understanding of the characteristics and structure of the data [16].

In the feature ranking process, we employed the correlation coefficient algorithm. The core of this algorithm lies in evaluating the linear relationships between features in the data, thereby identifying the features that have the most significant impact on the target variable. By calculating the correlation coefficients between each feature and the target variable, we are able to quantify the strength and direction of these relationships. This not only helps in understanding the extent to which different features influence the prediction results but also provides a solid foundation for subsequent feature selection. We then used an incremental feature selection strategy to progressively build feature subsets based on the importance of the features. Starting with the most important features, we gradually included less important features to generate feature subsets. These were then combined with a spectral clustering model, and the optimal feature subset was determined based on the results of five-fold cross-validation. The results were plotted in an iterative feature selection curve, as shown in Figure 1, where the horizontal axis represents the number of features and the vertical axis represents the silhouette coefficient score, with the highest point indicating the optimal feature subset.

From the graph, it can be seen that the feature subset comprising the top 29 features, ranked using the correlation coefficient algorithm, is the optimal feature subset.

### 5.3. Prediction performance of model

After selecting the optimal feature subset, we carried out comprehensive experiments on the dataset to identify the most accurate clustering algorithm using the selected feature set. We explored three different clustering algorithms, namely spectral clustering, agglomerative clustering, and K-means clustering, to build models. The performance of these models was assessed using metrics such as the ARI, NMI, Silhouette Score, CHI, and DBI. The results, after Jackknife cross-validation, are presented in Table 1. Additionally, we visualized the clustering results using principal component analysis, as shown in Figures 2, 3, and 4. From the results, the three clustering algorithms demonstrate varying strengths across different metrics. K-means exhibits the best overall performance, achieving the highest Mean ARI (0.7296) and Mean NMI (0.7393), indicating the strongest alignment with true labels and the highest level of shared information. This makes it the most effective algorithm for distinguishing samples. Spectral clustering follows closely, with

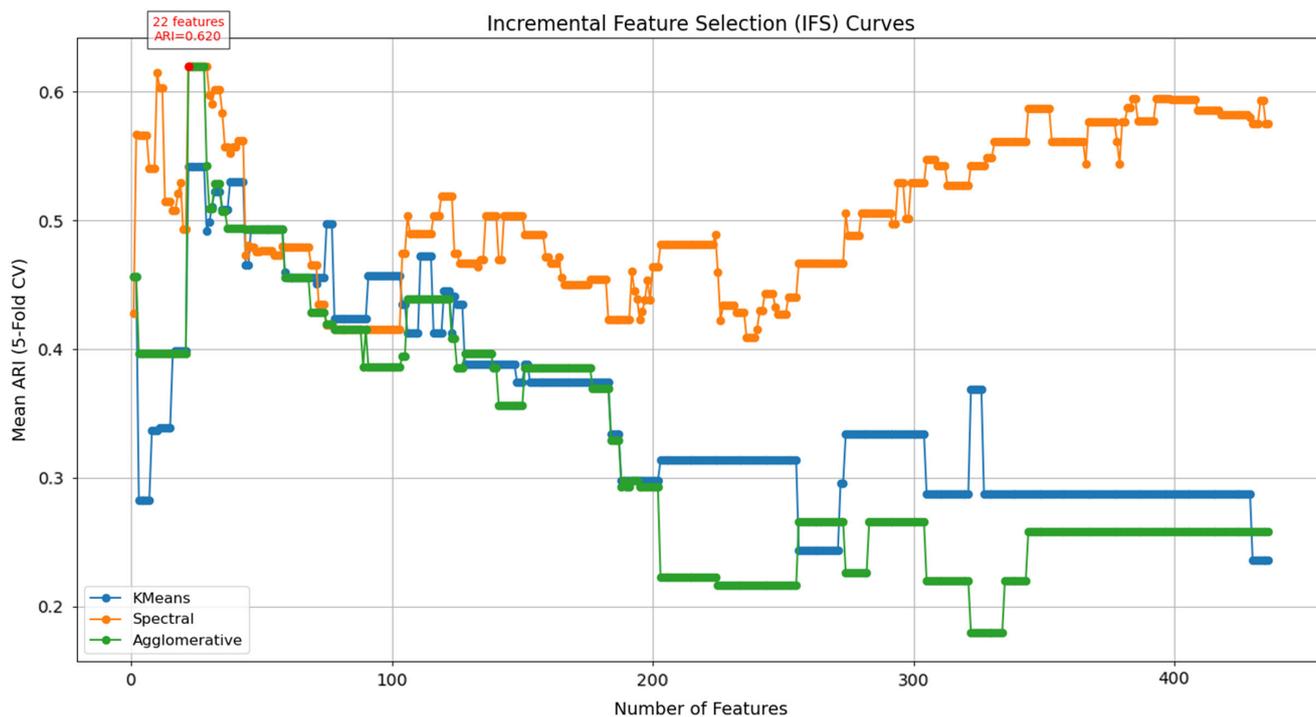


Figure 1. ARI for incremental feature subsets

Table 1. Results of Jackknife cross-validation under three clustering algorithms

	Spectral clustering	K-means	Agglomerative clustering
Mean ARI	0.7105	0.7296	0.6607
Mean NMI	0.7221	0.7393	0.7032
Mean Silhouette Score	0.3278	0.3286	0.3313
Mean CHI	34.1959	34.2127	28.9767
Mean DBI	1.1796	1.1651	1.2040

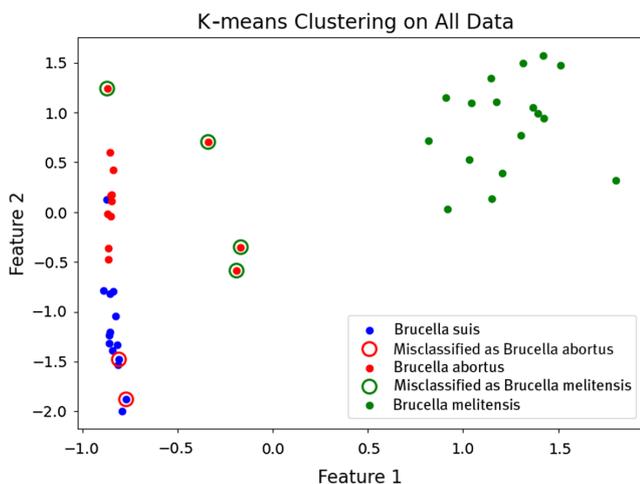


Figure 2. Visualization of clustering results under K-means clustering

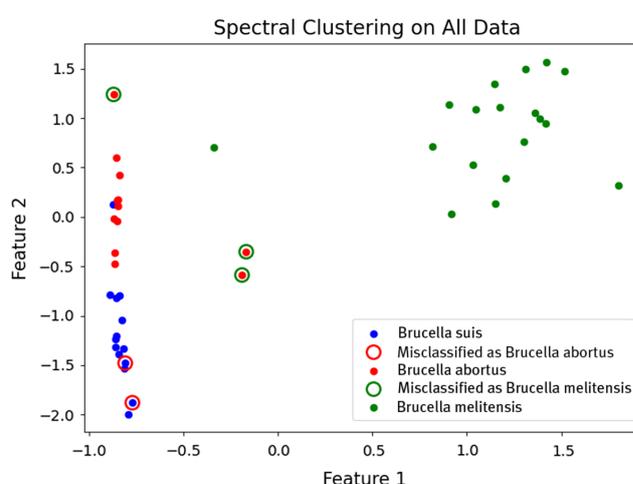
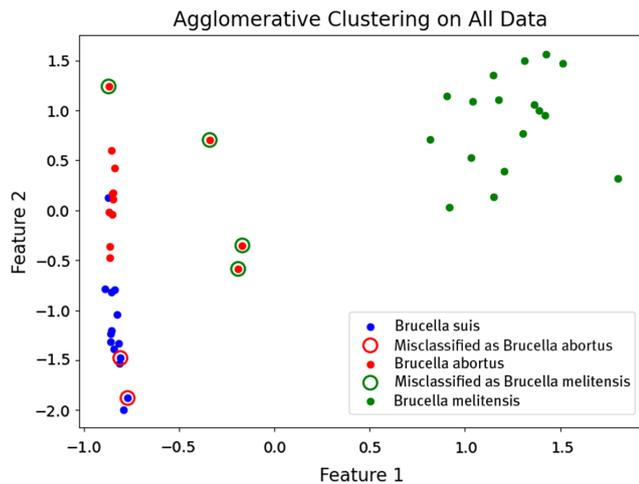


Figure 3. Visualization of clustering results under spectral clustering



**Figure 4. Visualization of clustering results under agglomerative clustering**

slightly lower ARI (0.7105) and NMI (0.7221), but it still performs well, particularly for handling complex, nonlinear data structures. Agglomerative clustering has the lowest ARI (0.6607) and NMI (0.7032), suggesting weaker correlation with true labels and less information overlap, resulting in inferior clustering quality compared to the other two methods. The Silhouette Scores are similar across all three, indicating comparable internal consistency. The CHI shows K-means is slightly superior, and its lowest DBI score further validates its advantage.

## 6. Educational Reform Strategies

### 6.1. Strategic framework for reform

Discrete mathematics is the study of fundamentally discrete mathematical structures [17], which is crucial for computer science, cryptography, algorithm design, etc. However, traditional teaching methods often fail to fully attract students or to illustrate the practical applications of these concepts.

As we are in an era that computational thinking and digital literacy are important, rethinking how we teach the course of discrete mathematics becomes imperative. Adopting advanced strategies in discrete mathematics teaching is crucial, because it helps students with vital knowledge and skills. This kind of integration helps create a well-rounded learning environment. The following is the plan to improve discrete mathematics education.

### 6.2. Curriculum integration with real-world applications

The key of educational reform in discrete mathematics curriculum is to combine discrete mathematics with real-world applications. The relationship between industry and courses is very important because the case studies and projects reflect the real applications of industry. Such as we apply discrete mathematics in clustering algorithms, specifically focusing on the traceability of Brucella strains, integrating case studies of clustering algorithms into the curriculum of a discrete mathematics course.

In the area of interdisciplinary projects, teachers can show how discrete mathematics applies in various fields, such as genetics [18] and evolution in biology [19], game theory in economics and social sciences. This approach can narrow the gap between disciplines and motivate students to solve problems across various domains by applying mathematical concepts. The benefit is that students can know the relevance of discrete mathematics beyond the classroom, increasing their understanding and interest of the subject.

### 6.3. Adoption of modern educational approaches

Using modern teaching methods in the discrete mathematics courses is the important to help student understanding and joining the course content [20]. Flipped classroom model is an effective way that shifts traditional learning by having students learn new content at home, often through online lectures or videos. As a result, classroom time is available for discussions and experiment. This transformation turns the classroom into a dynamic learning space where students actively apply concepts to solve problems and join in group projects. This kind of participation rises understanding of course content instead of passive learning. In our study cases, students can present their learned clustering algorithms to others, and they talk together and try to improve these algorithms in different ways. Such activities can exercise students' teamwork and communication skills.

### 6.4. Focus on skills beyond mathematics

In the discrete mathematics curriculum, students are suggested to focus on developing skills beyond traditional mathematics in order to fulfill the needs of the modern workforce. The problem-solving abilities is central to this kind of skills, which teaches students how to solve mathematical problems, deal with issues systematically, think critically, and apply innovative solutions. This type of expertise is useful in all fields and in everyday life, helping students to approach challenges from multiple perspectives and generate unique solutions.

It is also crucial to incorporate computational thinking into the curriculum, which includes programming activities and algorithmic thinking. The curriculum encourages students to get familiar with programming languages such as Python or Java, which improves their understandings of cases in discrete mathematics and help them practice programming skills to solve problems in the tech industry and beyond.

### 6.5. Professional development for teachers

The professional development of teachers is a critical component of discrete mathematics education reform because teaching quality has a significant impact on student engagement. This part of reform offers instructors with continual learning opportunities, such as workshops, seminars, and courses meant to keep them up to date on the latest educational technologies, innovative teaching practices, and breakthroughs in their profession.

Additionally, fostering a culture of peer cooperation through online forums and conferences allows teachers to share their resources and experiences, resulting in a learning community that extends beyond classrooms or institutions. This collaborative approach to professional development improves teachers' individual competencies and raises the overall quality of instruction within the subject. What's more, it can improve

student results and stimulate innovation in discrete mathematics education.

## 6.6. Implementing the strategy

Teachers, institutions, and the larger educational community must all commit to and collaborate on these ideas for them to be implemented successfully. Pilot projects can be launched to test new approaches, collect feedback, and make required changes. Furthermore, engaging students in the reform process allows them to provide feedback and ideas, which can considerably improve the curriculum's relevance and efficacy.

## 7. Discussion

Incorporating case studies of clustering algorithms into discrete mathematics courses is a key step in linking theoretical principles with practical applications. In particular, clustering algorithms have broad potential applications in computational biology, playing an important role in areas such as gene expression analysis, disease subtype classification, single-cell RNA sequencing, protein network module detection, epigenome integration, microbiome classification, and molecular dynamics analysis [21–23]. Their advantages in handling high-dimensional nonlinear data enable effective pattern discovery, thereby supporting precision medicine and systems biology research.

The key to making this teaching style interactive is to foster an environment in which active participation is encouraged and considered an essential component of the learning experience. This is accomplished by having students participate in activities such as group discussions, case studies, and simulated experiments [24]. These activities are intended to assist students to apply theoretical knowledge in practical situations. By immersing students in this way, abstract concepts of discrete mathematics are clarified while their practical application is readily demonstrated through concrete examples.

This teaching technique relies on providing students application scenarios of real-world datasets. This approach makes students to directly apply clustering techniques to specific problems [25], so it is useful to improve their knowledge of algorithms and demonstrate the wider mathematical ideas. Through such practical stage, students learn how to analyze data and receive a better understanding of problem-solving process. Students learn a wide range of useful skills by actively engaging the collaborative activities. Their analytical skills are improved when they solve difficult problems with datasets. Furthermore, the collaborative process improves teamwork and communication skills because students work together to overcome difficulties and share insights.

Moreover, teaching discrete mathematics through the use of clustering methods serves two purposes. Students can obtain a deeper understanding of mathematical theories and discover how these theories are applied across various fields directly. This awareness that mathematics is more than abstract concepts, but a versatile tool applicable across fields, which is critical in sparking students' interest in the subject. It opens up a variety of opportunities for them to apply their mathematical knowledge in their future work, preparing them both academically and professionally.

To summarize, by incorporating case studies of clustering algorithms into an interactive discrete mathematics course, teachers can provide a rich, engaging, and informative experience.

This technique not only improves students' understanding of mathematical ideas, but it also provides them with practical skills and prepares them for future employment opportunities. This approach reflects a holistic educational model in which learning is brought to life by active exploration, collaboration, and the application of knowledge to real-world circumstances, rather than being limited to textbooks.

## 8. Conclusion

In this paper, we comprehensively discuss the use of discrete mathematics in clustering methods, with a case study on the traceability of *Brucella* strains. Using three methods – spectral clustering, K-means clustering, and agglomerative clustering – we show how clustering algorithms work in practice. In this study, we first create a benchmark dataset with three different *Brucella* species: *Brucella melitensis*, *Brucella abortus*, and *Brucella suis*. We then utilize the correlation coefficient technique to rank over 436 features. Next, we use the spectral clustering algorithm to select 22 optimum features for the feature subset. Finally, we use three clustering algorithms to model the samples and analyze the results. This study not only successfully established a clustering model for the traceability of *Brucella* but also provided a general mathematical clustering algorithm workflow for solving similar problems, offering significant guidance for case-based teaching. Our discussion on advanced, forward-thinking strategies to reform discrete mathematics education highlights the critical importance of practical applications, modern pedagogies, technological integration, and skills development beyond traditional mathematics. It also underscores the value of continuous assessment and the professional development of teachers.

In the future, we plan to acquire more genomic data of various *Brucella* species to establish a more comprehensive traceability model. Additionally, we will apply the case study from this paper in actual teaching to further explore and optimize teaching methods. Through such research and practice, we aim to enhance the quality of teaching, enabling students to better understand and apply these complex mathematical and statistical concepts. This will also help students to master key skills in the fields of data science and bioinformatics, laying a solid foundation for their future careers.

## Funding Support

This study was supported by the National Natural Science Foundation of China (62271120, 62302065).

## Conflicts of Interest

The authors declare that they have no conflicts of interest to this work.

## Data Availability Statement

The data that support the findings of this study are openly available at <https://doi.org/10.3390/microorganisms10081658>, reference number [14]. No new data were generated.

## Author Contribution Statement

**Jiqing Gu:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

**Jing Hu:** Funding acquisition. **Ju Huang:** Software. **Hui Yang:** Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing – original draft, Writing – review & editing, Visualization, Supervision, Project administration, Funding acquisition.

## References

- [1] Greefrath, G., Siller, H. S., Vorhölter, K., & Kaiser, G. (2022). Mathematical modelling and discrete mathematics: Opportunities for modern mathematics teaching. *ZDM—Mathematics Education*, 54(4), 865–879. <https://doi.org/10.1007/s11858-022-01339-5>
- [2] Robeva, R., Davies, R., Hodge, T., & Enyedi, A. (2010). Mathematical biology modules based on modern molecular biology and modern discrete mathematics. *CBE—Life Sciences Education*, 9(3), 227–240. <https://doi.org/10.1187/cbe.10-03-0019>
- [3] Jungck, J. R., & Weisstein, A. E. (2013). Mathematics and evolutionary biology make bioinformatics education comprehensible. *Briefings in Bioinformatics*, 14(5), 599–609. <https://doi.org/10.1093/bib/bbt046>
- [4] Ellington, R., Wachira, J., & Nkwanta, A. (2010). RNA secondary structure prediction by using discrete mathematics: An interdisciplinary research experience for undergraduate students. *CBE—Life Sciences Education*, 9(3), 348–356. <https://doi.org/10.1187/cbe.10-03-0036>
- [5] Sali, M., de Maio, F., Tarantino, M., Garofolo, G., Tittarelli, M., Sacchini, L., . . . , & Adone, R. (2018). Rapid and safe one-step extraction method for the identification of *Brucella* strains at genus and species level by MALDI-TOF mass spectrometry. *PLoS One*, 13(6), e0197864. <https://doi.org/10.1371/journal.pone.0197864>
- [6] da Silva, D. A. V., Brendebach, H., Grütze, J., Dieckmann, R., Soares, R. M., de Lima, J. T. R., . . . , & Al Dahouk, S. (2020). MALDI-TOF MS and genomic analysis can make the difference in the clarification of canine brucellosis outbreaks. *Scientific Reports*, 10(1), 19246. <https://doi.org/10.1038/s41598-020-75960-3>
- [7] Ulitzsch, E., Khanna, S., Rhemtulla, M., & Domingue, B. W. (2023). A graph theory based similarity metric enables comparison of subpopulation psychometric networks. *Psychological Methods*. <https://doi.org/10.1037/met0000625>
- [8] Roizman, V., Jonckheere, M., & Pascal, F. (2024). A flexible EM-like clustering algorithm for noisy data. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 46(5), 2709–2721. <https://doi.org/10.1109/TPAMI.2023.3337195>
- [9] Chen, Y., Ye, W., & Li, D. (2023). Spectral clustering community detection algorithm based on point-wise mutual information graph kernel. *Entropy*, 25(12), 1617. <https://doi.org/10.3390/e25121617>
- [10] Liu, Y., Wang, N., Su, X., Zhao, T., Zhang, J., Geng, Y., . . . , & Huang, L. (2023). Classification of cognitive impairment in older adults based on brain functional state measurement data via hierarchical clustering analysis. *Frontiers in Aging Neuroscience*, 15, 1198481. <https://doi.org/10.3389/fnagi.2023.1198481>
- [11] Sim, J., Mani, K., Fazzari, M., Lin, J., Keller, M., Kitsis, E., . . . , & Jariwala, S. P. (2024). Using K-means clustering to identify physician clusters by electronic health record burden and efficiency. *Telemedicine and e-Health*, 30(2), 585–594. <https://doi.org/10.1089/tmj.2023.0167>
- [12] Zhang, W., Yue, Z., Ye, J., Xu, H., Wang, Y., Zhang, X., & Xi, L. (2022). Modulation format identification using the Calinski–Harabasz index. *Applied Optics*, 61(3), 851–857. <https://doi.org/10.1364/AO.448043>
- [13] Wang, G., Wang, Z., Chen, W., & Zhuang, J. (2006). Classification of surface EMG signals using optimal wavelet packet method based on Davies-Bouldin criterion. *Medical and Biological Engineering and Computing*, 44, 865–872. <https://doi.org/10.1007/s11517-006-0100-y>
- [14] Dematheis, F., Walter, M. C., Lang, D., Antwerpen, M., Scholz, H. C., Pfalzgraf, M. T., . . . , & Zange, S. (2022). Machine learning algorithms for classification of MALDI-TOF MS spectra from phylogenetically closely related species *Brucella melitensis*, *Brucella abortus* and *Brucella suis*. *Microorganisms*, 10(8), 1658. <https://doi.org/10.3390/microorganisms10081658>
- [15] Gholizadeh, M., Saeedi, R., Bagheri, A., & Paezi, M. (2024). Machine learning-based prediction of effluent total suspended solids in a wastewater treatment plant using different feature selection approaches: A comparative study. *Environmental Research*, 246, 118146. <https://doi.org/10.1016/j.envres.2024.118146>
- [16] Heinrich, F., Lange, T. M., Kircher, M., Ramzan, F., Schmitt, A. O., & Gültas, M. (2023). Exploring the potential of incremental feature selection to improve genomic prediction accuracy. *Genetics Selection Evolution*, 55(1), 78. <https://doi.org/10.1186/s12711-023-00853-8>
- [17] Hart, E. W., & Martin, W. G. (2018). Discrete mathematics is essential mathematics in a 21st century school curriculum. In E. W. Hart & J. Sandefur (Eds.), *Teaching and learning discrete mathematics worldwide: Curriculum and research* (pp. 3–19). Springer. [https://doi.org/10.1007/978-3-319-70308-4\\_1](https://doi.org/10.1007/978-3-319-70308-4_1)
- [18] Tenazinha, N., & Vinga, S. (2011). A survey on methods for modeling and analyzing integrated biological networks. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 8(4), 943–958. <https://doi.org/10.1109/TCBB.2010.117>
- [19] Robeva, R. S. (2015). *Algebraic and discrete mathematical methods for modern biology*. Netherlands: Elsevier.
- [20] Sandefur, J., Lockwood, E., Hart, E., & Greefrath, G. (2022). Teaching and learning discrete mathematics. *ZDM—Mathematics Education*, 54(4), 753–775. <https://doi.org/10.1007/s11858-022-01399-7>
- [21] Liu, C., Cheng, S., Ding, W., & Arcucci, R. (2025). Spectral cross-domain neural network with soft-adaptive threshold spectral enhancement. *IEEE Transactions on Neural Networks and Learning Systems*, 36(1), 692–703. <https://doi.org/10.1109/TNNLS.2023.3332217>
- [22] Liu, C., Cheng, S., Shi, M., Shah, A., Bai, W., & Arcucci, R. (2025). IMITATE: Clinical prior guided hierarchical vision-language pre-training. *IEEE Transactions on Medical Imaging*, 44(1), 519–529. <https://doi.org/10.1109/TMI.2024.3449690>
- [23] Wan, Z., Liu, C., Zhang, M., Fu, J., Wang, B., Cheng, S., . . . , & Arcucci, R. (2023). Med-UniC: Unifying cross-lingual medical vision-language pre-training by diminishing bias. In *Proceedings of the 37th International Conference on Neural*

*Information Processing Systems*, 56186–56197. <https://dl.acm.org/doi/10.5555/3666122.3668572>

- [24] Liu, Y. A., & Castellana, M. (2021). Discrete math with programming: A principled approach. In *Proceedings of the 52nd ACM Technical Symposium on Computer Science Education*, 1156–1162. <https://doi.org/10.1145/3408877.3432537>
- [25] Zhao, D., Li, H., Xu, A., & Song, T. (2022). Psychological mobilization of innovative teaching methods for students'

basic educational curriculum reform under deep learning. *Frontiers in Psychology*, 13, 843493. <https://doi.org/10.3389/fpsyg.2022.843493>

**How to Cite:** Gu, J., Hu, J., Huang, J., & Yang, H. (2025). Teaching Case Study in Discrete Mathematics: Application of Clustering Algorithms in Brucella Traceability Research. *Medinformatics*, 2(1), 49–56. <https://doi.org/10.47852/bonviewMEDIN52023060>