QUANTUM SPECTRAL CLUSTERING: AN HYBRID BASED QUANTUM KERNEL APPROACH.

QUANTUM SPECTRAL CLUSTERING: AN HYBRID BASED QUANTUM KERNEL APPROACH.

Mohamed Salah Jebali

Matricola: 7078487

School of Engineering - Master of Science Degree in Artificial Intelligence

Università degli Studi di Firenze

February 2024

Contents

ABSTRACT

This study investigates the integration of quantum computing techniques with spectral clustering algorithms, proposing and evaluating a hybrid approach that combines quantum kernels and feature maps with classical spectral clustering. Extensive experimentation on two datasets, including the ad_hoc dataset from Qiskit and a synthetic complex dataset, reveals that quantum spectral clustering, particularly utilizing the ZZFeatureMap, outperforms classical methods in identifying clusters, especially in datasets with non-linear separability and high complexity. However, computational trade-offs exist, with quantum methods showing higher accuracy on challenging datasets but requiring exponentially more computational time as dataset size increases. The findings underscore the current limitations of quantum simulation on classical hardware and highlight the need for advancements in quantum hardware development to realize the full potential of quantum algorithms. While promising, practical challenges such as scalability and computational efficiency remain, suggesting avenues for future research including scalability enhancement, adaptive feature maps, error mitigation strategies, and exploration of hybrid quantum-classical algorithms for more efficient utilization of quantum resources.

Keywords Quantum Machine Learning · Quantum Information Theory · Spectral Clustering · Quantum Spectral Clustering · Deep Learning · ZZFeatureMap · PualiFeatureMap · kernel methods

1 Introduction

The intersection of quantum computing and machine learning is a vibrant field of research that promises to redefine the boundaries of computational capabilities and data analysis. This exploration into **quantum spectral clustering** started from a simple question:
Can the principles of quantum computing enhance the way we group data, a foundational task in machine learning?

Clustering algorithms, which allow us to sift through data and discover natural groupings or patterns without predefined labels, are powerful tools for understanding complex datasets.
Among the various clustering algorithms, spectral clustering stands out for its ability to discern intricate structures within data, leveraging the eigenvalues and eigenvectors of similarity matrices to reveal the latent connections between data points.

Spectral clustering's elegance lies in its approach: it transforms the clustering problem into one of graph partitioning, where the goal is to find a cut through the graph that minimizes the disconnection between different clusters.
This method excels in situations where traditional clustering techniques fail, particularly with datasets that do not conform to spherical cluster shapes or when the clusters are unevenly sized or the points are not linearly separable.
However, spectral clustering is not without its challenges. It can be computationally intensive, particularly for large datasets, and its performance heavily depends on the choice of similarity matrix and the dimensionality reduction techniques employed.

The advent of quantum computing offers the possibility to find innovative solutions to these challenges. Quantum computers, with their ability to perform calculations in high-dimensional Hilbert spaces and exploit phenomena like superposition and entanglement, provide a new paradigm for processing information. The core hypothesis driving this project was that quantum computing could be harnessed to enhance spectral clustering, making it more efficient and capable of uncovering patterns that remain obscured to classical algorithms.

Inspiration for this research project was drawn from the groundbreaking work of Havlíček et al. Havlíček et al. (2019), titled **"Supervised learning with quantum-enhanced feature spaces"**. Their research demonstrated the potential of quantum computing to augment feature spaces in a way that significantly improves the performance of supervised learning models. This result laid down the groundwork for questioning how quantum-enhanced feature spaces could be applied beyond the realm of supervised learning, extending it into the domain of unsupervised learning tasks, such as clustering, and in particular spectral clustering.

This study focused on an hybrid strategy, centered on integrating quantum kernels with classical spectral clustering algorithms. Quantum kernels, through their capacity to compute the similarity between data points in an enhanced feature space, offered a promising way for capturing the subtle properties of complex datasets. I experimented with various quantum kernels, including the ZZFeatureMap, PauliFeatureMap, both from *Qiskit*, and custom feature maps, in search of the optimal configuration to improve clustering outcomes. I compared the quantum enhanced spectral clustering with the classic spectral clustering, both vanilla and PCA improved.

The empirical investigation of this hybrid approach spanned several datasets, each selected to test the efficacy of quantum spectral clustering under different conditions. The ad_hoc dataset, specifically designed to benchmark quantum machine learning algorithms, provided a test-bed for evaluating the potential of quantum approaches to outperform classical methods in identifying complex patterns.
A novel synthetically generated dataset, characterized by its non-linear separability and intricate structure, further challenged the algorithms, probing the limits of both classical and quantum spectral clustering.

Through quantitative and qualitative evaluations, including the use of PCA and t-SNE for preliminary data analysis, this exploration sheds light on the conditions under which quantum spectral clustering demonstrates

its strengths and the challenges that lie ahead. This study, from concept to experimental results, contributed to the ongoing research which aim is to bridge quantum computing and machine learning.

The following report is organized as follows: first, I will give an introduction to the spectral clustering algorithm in Section 2; then, I will explain the approach and the methodology adopted in Section 3; following this, I will present the experiments and the obtained results in Section 4; finally, I will offer my conclusions at Section 5.

2 Spectral Clustering Algorithm

Spectral clustering is a versatile algorithm capable of uncovering complex structures within data, structures that often pose challenges for traditional clustering techniques.
Unlike methods that rely on direct distance measures in the original feature space, spectral clustering operates in a transformed space. In particular, it operates in the spectral domain, leveraging the eigenvectors of the Laplacian of the similarity matrix. Here, the intrinsic geometries of data clusters become more discernible, making this approach particularly effective for identifying non-convex clusters. It excels in scenarios where the cluster structure is defined more by connectivity among data points than by proximity to a central centroid. In essence, spectral clustering transforms the clustering problem into a graph-cut problem, facilitating a more nuanced partitioning of the data based on the underlying graph structure.

As illustrated in Figure 1, spectral clustering demonstrates its strengths where classical clustering algorithms stumble. In the provided qualitative example, the data exhibit a non-convex behavior—a characteristic that conventional algorithms often fail to recognize, leading to suboptimal clustering. This situation highlights how spectral clustering effectively handles real-world data complexities, providing a strong alternative to traditional methods.

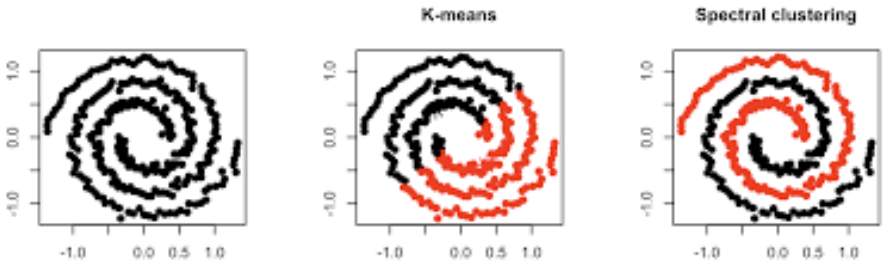

Figure 1: A qualitative comparison highlighting the advantage of spectral clustering in handling non-convex data structures, contrasted with the limitations of classical clustering algorithms.

Transformation into a Graph-cut Problem The elegance of spectral clustering lies in its foundational transformation of the clustering task into a graph partitioning problem. By representing the dataset as a graph, with nodes corresponding to data points and edges reflecting the similarity between pairs of points, spectral clustering seeks to partition this graph into subgraphs (clusters) in a way that minimizes the "cut" between them. This "cut" is defined as the sum of edge weights that would be removed by the partitioning. The goal is to achieve a partition where the connections (similarities) within each subgraph are strong (high similarity) while those between subgraphs are weak (low similarity), thus naturally grouping similar data points together.

Limitations of Classical Clustering Classical clustering algorithms, such as k-means, often assume that clusters are spherical and that all clusters have similar sizes and densities. These assumptions do not hold in many real-world scenarios, particularly when dealing with complex data structures. For datasets with non-linearly separable clusters, varying densities, or non-spherical shapes, classical clustering methods struggle to accurately identify the underlying patterns. This limitation highlights the need for more flexible clustering approaches capable of adapting to the data's intrinsic geometry, as spectral clustering does.

Understanding the Laplacian Matrix At the core of spectral clustering is the concept of the Laplacian matrix, a representation that captures the graph's structure. The Laplacian is derived from the similarity matrix, which quantifies the pairwise similarities between data points, and the degree matrix, which contains information about the number of connections each node has.
Mathematically, the unnormalized graph Laplacian is defined as $L = D - S$, where L is the Laplacian matrix, D is the diagonal degree matrix, and S is the similarity matrix.
The Laplacian matrix plays a crucial role in spectral clustering as its eigenvalues and eigenvectors are used to transform the data into a space where clusters can be more easily identified. This transformation is based on

the insight that the eigenvectors associated with the smallest non-zero eigenvalues of the Laplacian matrix contain essential information about the graph's connectivity and, consequently, about the natural grouping of the data.

The algorithm So, after having explained the crucial part of the algorithm, I will explain the steps followed by it. As introduced, starting by a set of points, we will transform it to a graph structure, by considering each point a node and the distances as edges. After computing the distances between points, we will create the degree matrix (D) and the similarity matrix (S) for computing the Laplacian matrix (L). We will then compute the first k eigenvectors of the Laplacian and apply the classical clustering algorithm on them. It is basically a shift of space, instead of clustering the points, we will cluster the eigenvectors of the Laplacian matrix. The algorithm is described in Algorithm 1 below:

Algorithm 1 Spectral Clustering Algorithm

Input: Data points $X = \{x_1, x_2, \ldots, x_n\}$, number of clusters k, similarity parameter σ
Output: Cluster labels for the data points
Construct the similarity matrix S using Gaussian kernel:
for $i, j \in \{1, \ldots, n\}, i \neq j$ **do**
    $S(i,j) = \exp\left(-\frac{\|x_i - x_j\|^2}{2\sigma^2}\right)$
end for
Compute the degree matrix D:
for $i \in \{1, \ldots, n\}$ **do**
    $D(i,i) = \sum_j S(i,j)$
end for
Compute the Laplacian matrix $L = D - S$
Perform eigenvalue decomposition on L to obtain its eigenvalues and eigenvectors
Select the k smallest eigenvectors $\{v_1, v_2, \ldots, v_k\}$ to form matrix $V \in \mathbb{R}^{n \times k}$
For each row i of V, treat it as a point y_i in $\mathbb{R}^k$
Apply k-means clustering to $\{y_1, y_2, \ldots, y_n\}$ to partition the points into k clusters
return Cluster labels obtained from k-means

Spectral clustering, through its unique use of the graph Laplacian and the subsequent eigenvalue decomposition, offers a powerful alternative to classical clustering methods, enabling the discovery of complex cluster structures that are otherwise difficult to detect. It is particularly able to identify cluster of arbitrary shapes and sizes, which leads a significant advantage over methods that assume spherical clusters, such as k-means. Moreover, it is particularly effective for datasets where the cluster structure is defined more by connectivity than by centroid proximity, allowing for more nuanced groupings.

Despite its advantages, spectral clustering comes with its own set of challenges, including the selection of an appropriate similarity measure, the determination of the optimal number of clusters, and computational considerations, especially for very large datasets. These factors necessitate a careful and informed application of spectral clustering to fully leverage its potential in uncovering the latent structures within data. In particular, its limitations can be divided in the following points:

- **Computational Complexity:** The most significant computational burden comes from the eigenvalue decomposition step, which can be prohibitively expensive for large datasets. The computational complexity of eigenvalue decomposition is generally $O(n^3)$ for dense matrix *n x n*, which can become prohibitively expensive for large datasets.
- **Sensitivity to Parameter Selection:** The performance of spectral clustering heavily depends on the choice of parameters, including the similarity measure and the number of clusters.
- **Handling of High-Dimensional Space:** Although spectral clustering excels in uncovering complex cluster structures, its efficacy can diminish in extremely high-dimensional spaces due to the curse of dimensionality.

3 Quantum Spectral Clustering: the proposed approach

Spectral clustering offers a sophisticated framework for identifying intricate cluster structures within datasets, transcending the limitations of traditional clustering methods. However, its computational intensity, particularly for large dataset, and the difficulty on high-dimensional datasets highlight the need for more efficient and accurate solutions. This is where the quantum approach comes into play.

Quantum computing, with its inherent ability to operate in high-dimensional Hilbert spaces and perform computations that exploit quantum parallelism, presents a promising avenue for overcoming the limitations of spectral clustering. By leveraging quantum-enhanced feature spaces and quantum algorithms, it may be possible to achieve more accurate clustering results, particularly for complex datasets that challenge classical computational paradigms.

3.1 The Hybrid Approach and Quantum Kernels

The proposed method for the quantum spectral clustering algorithm, adopted in this work, makes use of an hybrid approach, between classical and quantum algorithms. It strategically combines the strengths of classical machine learning techniques with the transformative potential of quantum computing.
The core methodology of the hybrid approach lies on the kernel function: the quantum computing element is leveraged in the kernel computation phase, while the remaining aspects of the spectral clustering algorithm remain grounded in classical computing techniques.
This strategy is devised to enhance the spectral clustering algorithm's capability to process complex, high-dimensional data, which often poses significant challenges to purely classical approaches

Delineating the Hybrid-Model In the hybrid model, the classical component leverages the spectral clustering algorithm, renowned for its effectiveness in identifying clusters through the spectral analysis of similarity matrices.
The quantum component introduces quantum kernels into the process, enabling the encoding of data into complex quantum states. This allows quantum kernels to evaluate data point similarities in a manner that captures the underlying structure of the data more effectively than classical methods, potentially leading to more accurate clustering results.

A kernel in machine learning is essentially a function that computes the similarity between pairs of data points in a feature space. Classical kernels achieve this by either directly working in the original feature space or transforming the data into a higher-dimensional space to make it easier for algorithms like *k-means* to identify clusters.
Mathematically, a classical kernel between points x_i and x_j is given by the kernel function $K(x_i, x_j)$, where the function K encapsulates the specific transformation and similarity measure.

Building upon the intuition of classical kernel methods, the concept of quantum kernels becomes more accessible. Initially, let's delve into the realm of a quantum feature space. Given a classical data vector $\mathbf{x}$, it is transformed into a quantum state $|\phi(\mathbf{x})\rangle$ through an encoding circuit $U(\mathbf{x})$, detailed as follows:

$$U(x)|0^{\otimes N}\rangle = |\phi(x)\rangle. \tag{1}$$

This process of encoding essentially serves as a quantum feature map, transitioning from classical data space to Hilbert space. Anchored in this conceptual framework, we introduce a quantum kernel function defined by the inner product of two quantum feature vectors within the Hilbert space, expressed as:

$$K_Q(x_i, x_j) = |\langle\phi(x_i)|\phi(x_j)\rangle|^2, \tag{2}$$

which further evolves into:

$$|\langle\phi(x_i)|\phi(x_j)\rangle|^2 = |\langle 0^{\otimes N}|U^\dagger(x_j)U(x_i)|0^{\otimes N}\rangle|^2. \tag{3}$$

In Fig. 2 it is shown the basic circuit that is able to approximate the quantum kernel function delineated in Eq. 3. This methodology, referred to as Quantum Kernel Estimation (QKE), paves the way for constructing quantum kernels.

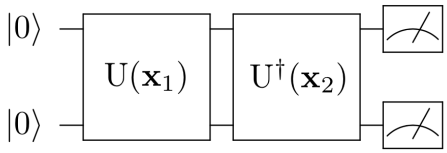

Figure 2: A Quantum Kernel Estimation circuit architecture.

The power of these quantum kernels lies in their capability to identify patterns that are difficult for classical methods, indicating a quantum advantage. This advantage comes from the fact that quantum kernels cannot be simulated classically, paving the way for new possibilities in pattern recognition beyond what classical methods can achieve.

Quantum kernels diverge from classical kernels by utilizing the principles of quantum mechanics to project data into a quantum feature space, enabling similarity measurements in a dimensional rich Hilbert space. This quantum projection is achieved through quantum feature maps, which encode classical data into quantum states, as shown in Fig. 3

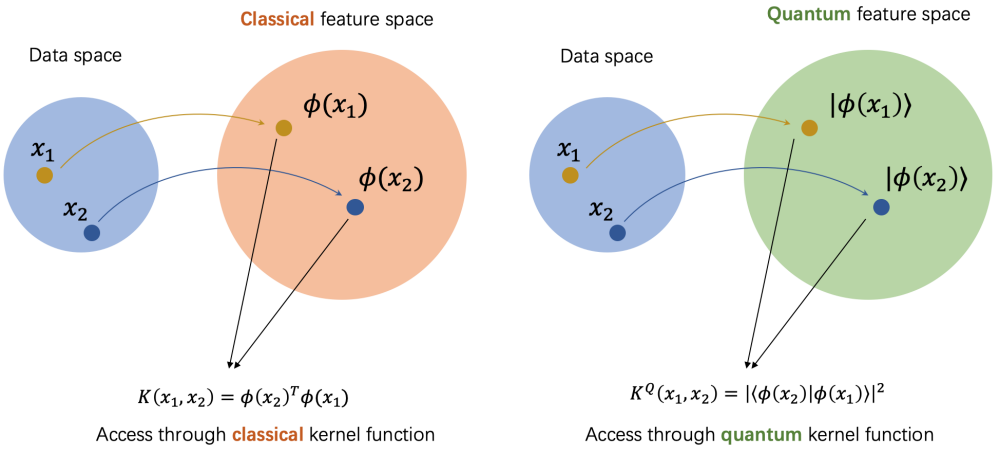

Figure 3: A comparative illustration highlighting the fundamental differences between quantum kernels and classical kernels.

Employed Quantum Feature Maps and Quantum Kernel This study investigates three distinct approaches to quantum kernels, each with unique properties and potential benefits for the spectral clustering task. I tried three different strategies for the quantum feature maps, as follows:

- **ZZFeatureMap:** Employs entangling gates and Pauli-Z rotations to encode data into quantum states, emphasizing the capture of second-order interactions between features.
- **PauliFeatureMap:** Utilizes a broader set of Pauli gates (X, Y, ZZ) to encode data, allowing for a richer exploration of feature relationships within the quantum feature space.
- **Custom Feature Map:** Offers a tailored approach to quantum state preparation, enabling optimizations specific to the dataset's characteristics or the clustering objectives.

Each quantum feature map has different properties and behaviors and they have been all used with the same quantum kernel, the **Fidelity Qauntum Kernel** from Qiskit.

These quantum kernel approaches represent a pivotal component of the hybrid spectral clustering model, promising to unlock new possibilities in the analysis of complex datasets. The subsequent subsections will provide a detailed examination of the quantum kernel used and each approach for the feature maps, including their mathematical foundations, circuit implementations, and the specific advantages they bring to the quantum spectral clustering framework.

3.2 Fidelity Quantum Kernel

In the quantum-enhanced spectral clustering studied here, the Fidelity Quantum Kernel is crucial across all tested feature maps. This part explains how the Fidelity Quantum Kernel works and how we use it in our quantum machine learning framework.

3.2.1 Operational Mechanics

The Fidelity Quantum Kernel computes the similarity between data points transformed into quantum states by different quantum feature maps. It is based on the concept of state fidelity, which measures the closeness between two quantum states. Here's a breakdown of its operational mechanics.

State Preparation and Encoding Each classical data point is first encoded into a quantum state through a specific quantum feature map, such as ZZFeatureMap, PauliFeatureMap, or a custom feature map. This encoding translates classical information into the quantum domain, preparing it for quantum processing.

Quantum Circuit Execution A quantum circuit is constructed to implement the feature map and facilitate the computation of fidelity between the quantum states corresponding to pairs of data points. This step involves applying the quantum feature map to prepare the states and additional operations to measure the overlap between these states.

Fidelity Measurement The core of the Fidelity Quantum Kernel's operation is the measurement of fidelity between quantum states. For two data points encoded as quantum states $|\phi(x_i)\rangle$ and $|\phi(x_j)\rangle$, their fidelity is calculated using the formula:

$$F(x_i, x_j) = |\langle\psi(x_i)|\psi(x_j)\rangle|^2, \tag{4}$$

where $|\langle\psi(x_i)|\psi(x_j)\rangle|^2$ represents the squared magnitude of the inner product between the two states, quantifying their similarity.

Kernel Matrix Construction By systematically computing the fidelity for every pair of data points in the dataset, a kernel matrix is constructed. This matrix, reflective of the quantum-enhanced similarities between data points, is then utilized in the spectral clustering algorithm to perform data segmentation based on the

intricate patterns revealed in the quantum feature space.

The Fidelity Quantum Kernel thus enables the leveraging of quantum computational advantages for machine learning, particularly in tasks requiring the analysis of complex, high-dimensional data. Its application through various feature maps offers a flexible approach to uncovering data relationships that may be less apparent or inaccessible to classical computational methods.

3.3 ZZFeatureaMap

The ZZFeatureMap is a quantum feature map used to encode classical data into quantum states for machine learning applications, including spectral clustering. It is specifically designed to exploit quantum mechanics to process data in ways that classical systems cannot, by mapping input features onto a quantum system's state space. This encoding is pivotal for implementing quantum kernels, which measure the similarity between data points in a quantum-enhanced feature space.

The ZZFeatureMap is part of a broader category of feature maps known as Pauli Feature Maps, which utilize Pauli gates (X, Y, Z) to perform the encoding. Specifically, the ZZFeatureMap employs Z gates and controlled-Z operations to introduce correlations between qubits based on the input data, facilitating the creation of a high-dimensional feature space that captures complex patterns in the data. This method allows for the exploration of data relationships that are difficult to discern using traditional linear algebra-based techniques, thereby enhancing the efficacy of algorithms like spectral clustering.

Mathematical Foundation The ZZFeatureMap operates by applying a series of quantum gates to encode classical data into a quantum state. Specifically, it utilizes Z rotations and controlled-Z CZ gates to create entanglements and encode feature interactions. The mathematical representation of the ZZFeatureMap acting on a set of features $(x_1, \ldots, x_n) \in \mathbb{R}^n$ can be expressed as a sequence of unitary operations:

$$U_{ZZ}(x) = exp(i \sum_{k<j} x_j x_k Z_j \otimes Z_k) \tag{5}$$

This indicates that for each pair of features x_j and x_k, the map applies a rotation governed by their product, modulating the entanglement between the corresponding qubits via the $Z_j \otimes Z_k$ operation. The Z rotation gates introduce phase shifts that encode the magnitude of the features, while the CZ gates entangle qubits, capturing the inter-feature relationships.

Application in Spectral Clustering The application of the ZZFeatureMap within the spectral clustering framework is a multi-step process that involves encoding classical data into quantum states, computing a quantum kernel matrix, and then using this matrix within a classical spectral clustering algorithm to identify clusters. The process is formalized as follows:

Algorithm 2 Application of ZZFeatureMap in Spectral Clustering

Require: Classical data points $X = \{x_1, x_2, \ldots, x_n\}$, feature dimension d, number of repetitions $reps$
Ensure: Cluster labels for the data points

1: Initialize ZZFeatureMap with feature dimension d and repetitions $reps$ ▷ Prepares the quantum feature map
2: Define quantum kernel using FidelityQuantumKernel with ZZFeatureMap ▷ Sets up the kernel with quantum features
3: Compute kernel matrix K for all pairs in X ▷ Generates the similarity matrix
4: Apply spectral clustering to K to obtain clusters ▷ Performs clustering on the quantum-enhanced similarities
5: **return** Cluster labels from spectral clustering

This algorithm captures the integration of quantum computing techniques, specifically through the use of the ZZFeatureMap, into the process of spectral clustering. By encoding classical data into quantum states and

utilizing the resulting quantum feature space for similarity measurements, we enhance the clustering process beyond the capabilities of classical computing alone.

Visualizing the ZZFeatureMap To fully appreciate and understand the ZZFeatureMap, we can have a look to the visual representation of its corresponding quantum circuit. This visualization helps understand the process of encoding classical data into quantum states and also illustrate the quantum mechanical interactions involved during this transformation.

The ZZFeatureMap employs a series of quantum gates to intricately encode input data onto qubits, the basic units of quantum information. Specifically, it utilizes Z and the controlled-Z (CZ) gates, as previously said, to introduce phase shifts and entanglements that correlate directly with the features of the input data. These operations are the fundamentals for capturing the nonlinear interactions between features.

In Fig. 4 we can observe an image of the quantum circuit generated by the ZZFeatureMap for 3 qubits and 1 repetition and linear entanglement, where ϕ is a classical non-linear function, which behave like $\phi(x) = x$ and $\phi(x, y) = (\pi - x)(\pi - y)$.

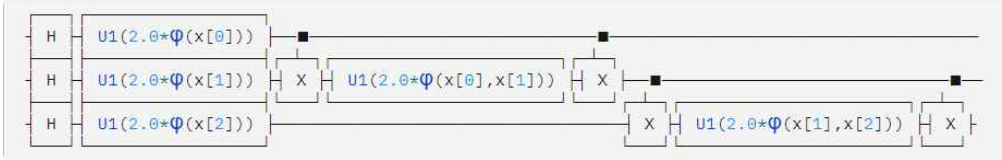

Figure 4: Quantum circuit representation of the ZZFeatureMap for 3 qubits and 1 repetition and linear entanglement, from Qiskit.

3.4 PauliFeatureMap

The PauliFeatureMap extends the concept of encoding classical data into quantum states by incorporating a broader set of quantum gates, specifically the Pauli gates (X, Y, Z). The use of Pauli gates allows for a diverse and rich encoding of data features into a quantum system's state space, facilitating a nuanced implementation of quantum kernels for similarity measurement.

Mathematical Foundation The PauliFeatureMap is designed to encode classical data into quantum states through the application of Pauli gates. Given a vector of input data $(x_1, \ldots, x_n) \in \mathbb{R}^n$, where n is the feature dimension, it transforms them as:

$$U_{\phi(\vec{x})} = exp(i \sum_{S \in \mathcal{I}} \phi_S(\vec{x}) \prod_{i \in S} P_i) \tag{6}$$

where S indicates a set of qubit indices that describes the connections in the feature map, $\mathcal{I}$ is a set containing all these index sets while $P_i \in \{X, Y, Z\}$. The data-mapping function ϕ_S behaves as:

$$\phi_S(\vec{x}) = \begin{cases} x_i, & \text{if } S = i, \\ \prod_{j \in S}(\pi - x_j), & \text{if } |S| > 1. \end{cases} \tag{7}$$

The diversity of Pauli gates used allows for a multifaceted encoding of feature interactions, capturing more complex data relationships than mappings using a single type of gate.

Application in Spectral Clustering For spectral clustering, the PauliFeatureMap is employed to transform classical data vectors into quantum states, which are then processed through a quantum kernel—specifically, the FidelityQuantumKernel—to compute the similarity matrix essential for clustering. Following, there is a structured approach to applying the PauliFeatureMap in the spectral clustering task:

Algorithm 3 Applying PauliFeatureMap in Spectral Clustering

Require: Classical data points $X = x_1, x_2, \ldots, x_n$, feature dimension d, number of repetitions $reps$
Ensure: Cluster labels for the data points

1: Initialize PauliFeatureMap with feature dimension d, repetitions $reps$, and specified Pauli gates ▷ Prepares the quantum feature map
2: Define the quantum kernel using FidelityQuantumKernel with the PauliFeatureMap ▷ Sets up the kernel with quantum features
3: Compute the kernel matrix K by evaluating the quantum kernel for all pairs in X ▷ Generates the similarity matrix
4: Apply spectral clustering to K to obtain clusters ▷ Performs clustering on the quantum-enhanced similarities
5: **return** Cluster labels from spectral clustering

Visualizing the PauliFeatureMap As for ZZFeatureMap, a visual representation of the PauliFeatureMap circuit can be useful for undesrtanding how classical data is econded into quantum states, through the application of Pauli gates.

Figure 5 showcases a circuit example featuring single-qubit Z rotations and two-qubit YY interactions across all pairs of qubits.

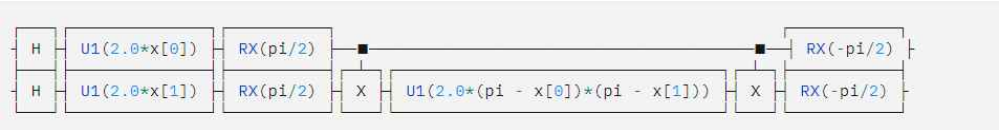

Figure 5: Quantum circuit representation of the PauliFeatureMap, illustrating the application of Pauli gates for data encoding: single-qubit Z rotations and two-qubit YY interactions across all pairs of qubits, from Qiskit.

While specific circuit diagrams depend on the chosen Pauli gates and the feature map's configuration, the principle remains consistent: Pauli gates introduce phase shifts and quantum entanglements reflective of the input data's features

3.5 Custom Feature Map

The idea of a Custom Feature Map in quantum machine learning allows us to adjust quantum state preparation to the unique characteristics of the dataset. Unlike preset feature maps like ZZFeatureMap or PauliFeatureMap, a custom feature map lets us design quantum circuits that encode data in ways that specifically improve machine learning models, such as spectral clustering.

Operational Overview Custom feature maps are constructed by explicitly defining the sequence of quantum gates that manipulate qubits to encode classical data into quantum states. This approach provides the freedom to experiment with different combinations of quantum gates to discover the most effective way to represent data within the quantum feature space.

In my case, I tried a custom feature map that leverages Hadamard gates to create superposition and RZ-gates to encode the actual data feature into the phase of each qubit. Controlled-NOT (CNOT) gates are used to introduce entanglement between qubits, enhancing the representation of feature correlations within the quantum state. The custom feature map can be mathematically hint at by the sequence of operations applied to qubits:

$$U_{\text{custom}}(x) = H^{\otimes n} \cdot \bigotimes_{i=1}^{n} RZ(x_i) \cdot \text{CNOT}_{\text{chain}} \tag{8}$$

where $H^{\otimes n}$ denotes the application of Hadamard gates across all qubits, creating a uniform superposition, $RZ(x_i)$ represents the rotation about the Z-axis by an angle proportional to the feature value x_i and CNOT_{chain} symbolizes a sequence of CNOT gates for entangling qubits.

Application in Spectral Clustering The application of a custom feature map within the spectral clustering process involves several key steps, similar to the procedure for predefined feature maps but with the flexibility of the custom encoding strategy:

Algorithm 4 Applying Custom Feature Map in Spectral Clustering

Require: Classical data points $X = \{x_1, x_2, \ldots, x_n\}$, feature dimension d, number of repetitions $reps$
Ensure: Cluster labels for the data points

1: Define a custom feature map with specified quantum gates and parameters ▷ Customizes the quantum state preparation
2: Initialize the quantum kernel with the custom feature map ▷ Sets up the kernel to measure quantum-enhanced similarities
3: Compute the kernel matrix K for all data point pairs in X ▷ Evaluates similarities in the quantum feature space
4: Use spectral clustering on K to determine data clusters ▷ Clusters data based on quantum state similarities
5: **return** Cluster labels derived from spectral clustering

After delving into the theoretical framework and algorithmic structure of three different quantum feature maps—ZZFeatureMap, PauliFeatureMap, and a custom feature map—we now turn to practical testing. This transition from theory to application is essential for confirming the effectiveness of quantum spectral clustering. In the forthcoming experiments section, I will conduct evaluations to show how these quantum-enhanced methods perform across different datasets. The aim is to highlight the concrete benefits and possible enhancements that quantum computing offers to machine learning, particularly in spectral clustering.

4 Experiments

This section presents my experimental approach, designed to test the effectiveness of hybrid quantum methods in spectral clustering compared to traditional classical methods. The main goal is to see if quantum algorithms can outperform classical ones, especially with complex datasets.

To conduct a thorough and detailed evaluation, I explored a hybrid quantum approach tailored for spectral clustering. Experiments were set up to directly compare quantum and classical approaches, aiming to find out which is better for handling challenging datasets. I focused on three quantum kernel-based methods—ZZFeatureMap, PauliFeatureMap, and Custom FeatureMap—and compared them with two classical spectral clustering methods: the standard approach and PCA-reduced approach.

I carefully selected a range of datasets for this study, covering both synthetic and complex already existing data, each with its unique challenges:

- **ad_hoc dataset:** from Qiskit, which has a complex structure.
- **synthetic generated dataset:** A synthetic 3D generated dataset, designed to be non-linearly separable with complex shapes.

Furthermore, I conducted multiple runs with the best performing models to compare how computational time varies with the number of data points. This helped to understand the efficiency and scalability trade-offs between quantum and classical methods.

In summary, this overview introduces the framework and objectives of my comparative study between hybrid quantum approaches and classical methods in spectral clustering. The subsequent sections will detail the experimental setup, describe the datasets utilized, elucidate the methodologies employed for both quantum and classical models, and present the findings of the experiments. I will also discuss the computational trade-offs identified through multi-run experiments, highlighting the efficiency and scalability of the selected models. This structured exploration aims to rigorously assess the capabilities and advantages of quantum-enhanced machine learning algorithms in addressing complex clustering tasks.

4.1 Experiment Setup

The experimental framework integrated both classical computing libraries and quantum computing frameworks to evaluate the performance of spectral clustering algorithms. The setup involved:

- **Qiskit version 0.46.0**: Serves as the foundational framework for constructing and executing quantum circuits.
- **Qiskit Machine Learning version 0.7.1**: Provides a suite of tools specifically designed for implementing quantum machine learning algorithms.
- **Qiskit Algorithms version 0.2.2**: Offers access to a variety of quantum algorithms essential for the experiments.
- **Python version 3.10.12**: The primary programming language used for developing the algorithms and performing data analysis.
- **Scikit-learn version 1.2.2**: Utilized for classical machine learning tasks, particularly for clustering and dimensionality reduction.
- **Quantum Simulator: Aer simulator from Qiskit-Aer package**: Employed for simulating quantum circuits.

Furthermore, to facilitate transparency and reproducibility, the complete codebase used for conducting these experiments is made available online. The code repository includes detailed instructions for setup and execution, ensuring that researchers can accurately reproduce the experimental environment and analyses.

For access to the source code, please visit the following link: Notebook for Quantum Spectral Clustering.

4.2 Dataset

In this study, I rigorously evaluate spectral clustering algorithms, with a focus on comparing classical and hybrid quantum approaches across a diverse set of datasets. The selection aims to provide a comprehensive assessment of the algorithms' performance under various conditions.

- **Quantum-Specific Datasets:** These datasets are designed to leverage the unique capabilities of quantum computing, such as high-dimensional feature spaces. They serve as a benchmark to highlight the potential advantages of quantum-enhanced clustering methods over classical approaches.
- **Synthetic Datasets:** Engineered to include challenging characteristics such as non-linear separability and complex geometrical configurations, synthetic datasets allow for controlled experimentation and assessment of the algorithms' robustness and adaptability.

To analyze the data distribution, complexity, and intrinsic structures within these datasets, advanced visualization techniques were employed. Principal Component Analysis (PCA) and t-Distributed Stochastic Neighbor Embedding (t-SNE) were utilized to reduce the dimensionality of the datasets, thereby facilitating an in-depth exploration of their underlying clustering patterns and structural nuances.

PCA provided insights into the global variance structure of the datasets, transforming the data into a lower-dimensional space that retains the most significant variance. Conversely, t-SNE focused on maintaining local data relationships in the dimensionality-reduced space, offering a granular view of the clustering dynamics. Together, these techniques form a comprehensive analytical framework for evaluating the datasets, setting the stage for a detailed examination of the clustering algorithms' efficacy across a spectrum of challenging scenarios.

4.2.1 Ad Hoc Dataset

Description The ad_hoc dataset is characterized by its colorful and complex structure, revealing the intricacy and non-linearity of the data points. It comprises two distinct classes: Class A and Class B, differentiated by markers (triangles and squares) and colors (yellow and red, respectively). This dataset poses a challenging clustering task due to the overlap between classes, the non-uniform distribution of data points, and the presence of complex geometrical patterns. The non-linear cluster boundaries further complicate the task, underscoring the limitations of traditional centroid-based clustering methods and highlighting the potential for quantum-enhanced algorithms to provide more nuanced and effective clustering solutions.

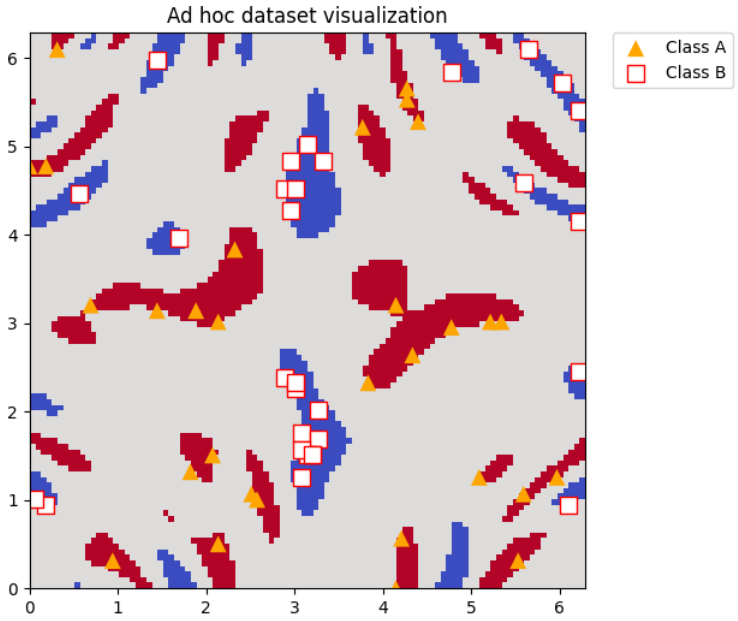

Figure 6: Visualization of the Ad Hoc Dataset

PCA Analysis The PCA reduction of the ad_hoc dataset reveals a complex data topology. Notably, the scatter of points along the principal components suggests significant overlap, indicating the potential non-linearity of class boundaries. The absence of distinct color blocks within the PCA plot further highlights the intricate nature of the dataset, where class transitions are gradual rather than discrete. Variations in point density across the principal components may hint at subclusters, underlining the challenge of identifying clear separations. This complexity points towards the necessity for more sophisticated clustering techniques that can navigate the nuanced structures inherent in the data.

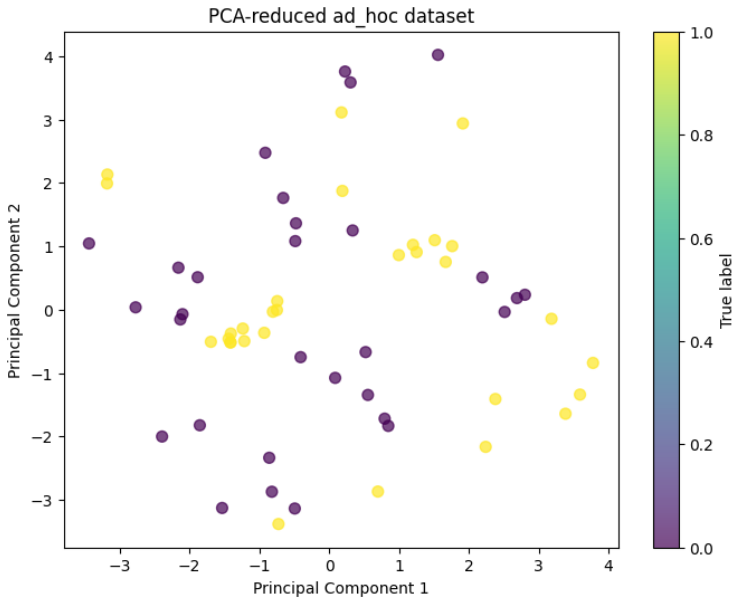

Figure 7: PCA of the Ad Hoc Dataset

T-SNE Analysis The t-SNE visualization of the ad_hoc dataset underscores the presence of local structures and overlapping regions between the two classes. The technique's focus on local neighborhood preservation is evident, suggesting clusters where data points are grouped based on class similarity, yet without a clear global separation. The plot reveals variations in data density and implies that the true class boundaries may be more nuanced than what is depicted in the reduced-dimensional space. This complexity corroborates the need for advanced clustering algorithms capable of interpreting such intricate data relationships.

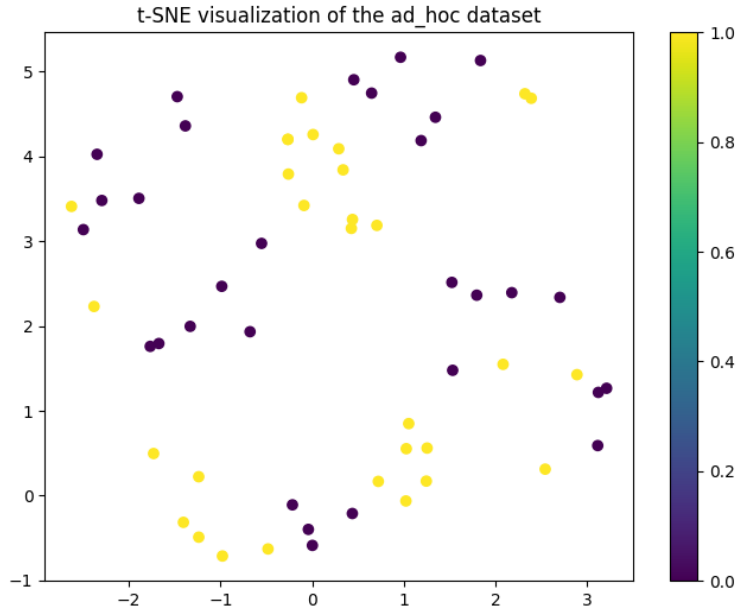

Figure 8: t-SNE Visualization of the Ad Hoc Dataset

This thorough examination of the ad_hoc dataset, using both PCA and t-SNE techniques, highlights its complexity and the potential difficulties it poses for clustering algorithms. The visualizations from these

techniques help us understand the dataset's intricate structures and serve as a benchmark for evaluating how well quantum spectral clustering handles it compared to classical methods.

4.2.2 Synthetic Dataset

Description The synthetic dataset, designed to emulate complex real-world data, is generated using the Swiss roll model, known for its three-dimensional, manifold-like structure. This dataset is further complicated by introducing a non-linear transformation to the 'z' dimension, replacing it with sine values to increase the dataset's intricacy. The visualization in a 3D plot showcases a twisted, convoluted form with a color gradient that reflects the manifold's curvature, emphasizing the necessity for advanced clustering techniques. The modified Swiss roll thus serves as a rigorous testbed to evaluate the potential of quantum spectral clustering against the limitations of classical linear methods.

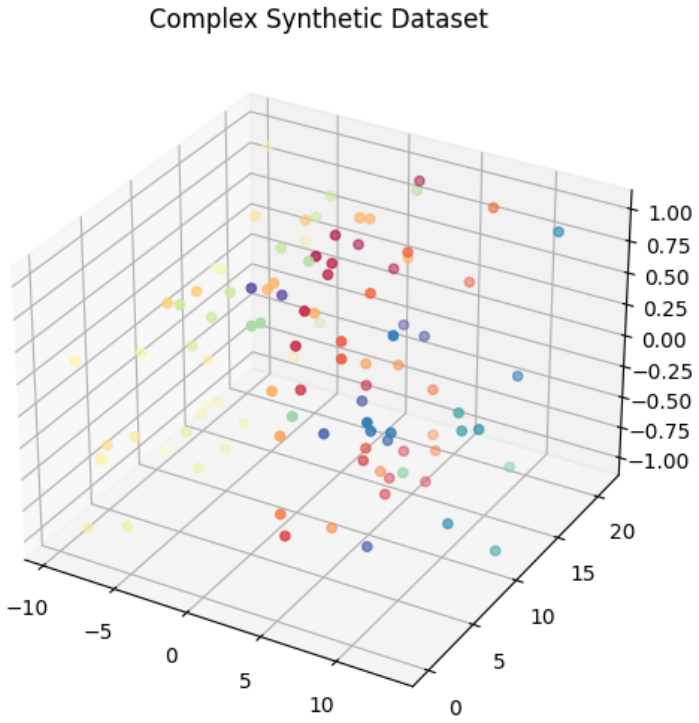

Figure 9: Three-dimensional Visualization of the Complex Synthetic Dataset

This synthetic dataset's complexity, characterized by its non-linear separability and intricate geometric structure, challenges traditional clustering algorithms and presents an ideal scenario to showcase the capabilities of spectral spectral clustering.

PCA Analysis Principal Component Analysis (PCA) on the synthetic dataset reduces its complexity from three dimensions to two, aiming to retain the most significant variances. The PCA plot indicates a wide spread of data points, suggesting a rich, continuous distribution without distinct clusters. The observed color gradient implies an inherent structure, hinting at the manifold's original shape. The PCA findings pose a challenge for clustering due to the absence of clear group boundaries, thus presenting a prime opportunity for quantum spectral clustering to demonstrate its advantage in capturing complex, high-dimensional relationships.

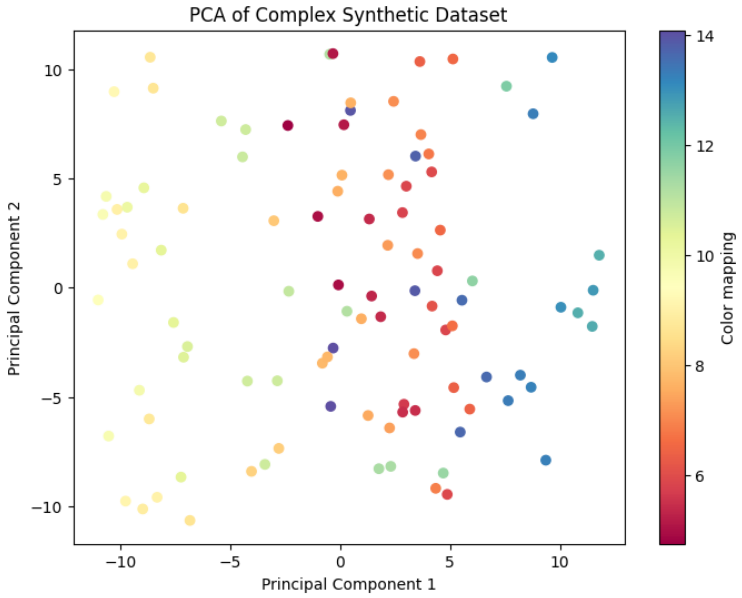

Figure 10: PCA Analysis of the Complex Synthetic Dataset.

The PCA reduction underscores the intricate nature of the dataset and sets the stage for assessing the quantum spectral clustering's ability to discern the subtle structures that classical PCA may not fully unveil.

T-SNE Analysis The t-SNE visualization of the synthetic dataset emphasizes its local clustering patterns, revealing groupings that closely match the dataset's color-coded structure. Unlike PCA, t-SNE shows clearer clusters, indicating its effectiveness in uncovering non-linear relationships. This ability to highlight clusters and dense areas underscores how quantum spectral clustering can delve deeper into these local patterns, providing a more thorough understanding of the dataset's complex geometry.

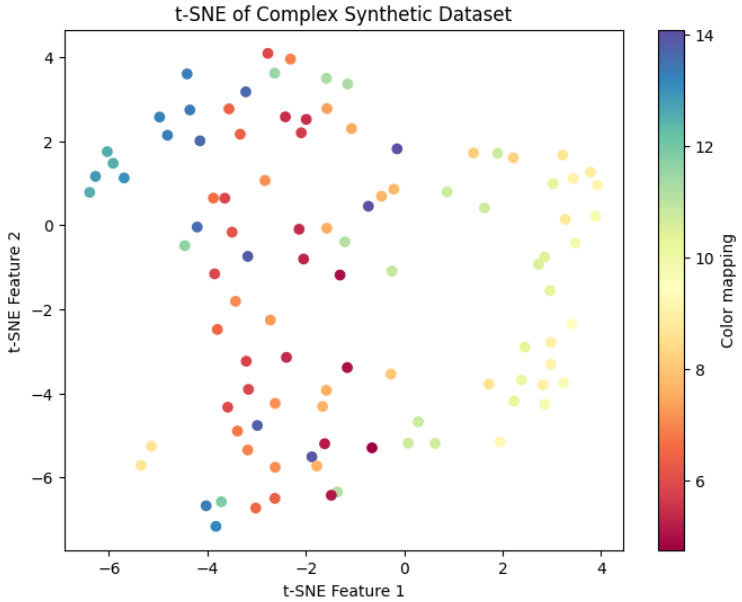

Figure 11: t-SNE Visualization of the Complex Synthetic Dataset

t-SNE's detailed portrayal of the dataset's topology is indicative of the diverse structures within, offering a strong case for the quantum approach's exploration to discern subtle, potentially quantum-advantaged patterns.

4.3 Experiments on Ad Hoc Dataset

In this Section, I will describe the experiments done on the ad_hoc dataset, which aim to understand if the quantum approach bring an effective advantage to the spectral clustering algorithm. Although I have conducted experiments with several architectures on different datasets, since the quantum circuit are dataset dependent, I have decided to separate the experiments based on the dataset. In this Section, the quantum circuits for this particular dataset will be shown, then quantitative results and qualitative results will be commented.

4.3.1 Quantum Architecture Description

In this section, I describe the quantum circuit architectures used for the experiments on the dataset. I will highlight the choice of hyperparameters, such as the number of qubits and repetitions (reps), for each type of quantum feature map (ZZFeatureMap, PauliFeatureMap, CustomFeatureMap). For each feature map, there is a figure of the quantum circuit to visually support the textual description.

ZZ Feature Map Quantum Circuit For the ad_hoc dataset, the ZZFeatureMap is configured with 2 qubits to match the number of classes in the dataset, with the circuit repeated twice to deepen the feature encoding. At the start, both qubits are passed through Hadamard gates (H), creating a superposition state that allows for the encoding of classical data into the quantum state space. After superposition, phase rotation gates (P) are applied to each qubit, with the rotations determined by the classical data points, effectively encoding the classical data into the quantum state. Subsequent controlled-Z (CZ) gates entangle the qubits, crucial for capturing correlations between different data features. Interspersed with CZ gates are additional phase rotations, further shaping the quantum state based on the classical features. The circuit ends with more phase rotations and a final round of entangling CZ gates, ensuring thorough integration of the data into the quantum state.

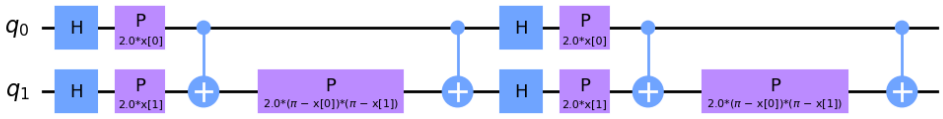

Figure 12: Quantum circuit representation of the ZZFeatureMap for the ad_hoc dataset, illustrating the initial Hadamard gates for state preparation, followed by parameterized phase rotations and entangling controlled-Z gates, repeated over two cycles to encode the classical data into a quantum feature space.

Pauli Feature Map Quantum Circuit

Pauli Feature Map Quantum Circuit For the ad_hoc dataset, the PauliFeatureMap is configured similarly with 2 qubits to align with the dataset's two classes, and the feature map circuit is executed in two repetitions. The circuit starts with Hadamard gates to create an equal superposition state across the qubits, setting the stage for data encoding. Phase rotation gates (P) follow, with each qubit's rotation parameterized by the corresponding classical feature to be encoded. Rotations around the x-axis (Rx gates) are then applied at $\pi/2$ and $-\pi/2$, introducing additional variability and complexity into the state preparation based on the classical data.

Controlled-phase (CP) gates are used to create a series of entanglements between the qubits, pivotal for capturing the high-dimensional interactions of the features. Subsequent alternating sequences of Hadamard,

phase rotations, and Rx gates further manipulate the quantum state, allowing for a richer encoding of the classical information into the quantum domain.

The PauliFeatureMap's quantum circuit design concludes with a final series of phase rotations and entanglement gates.

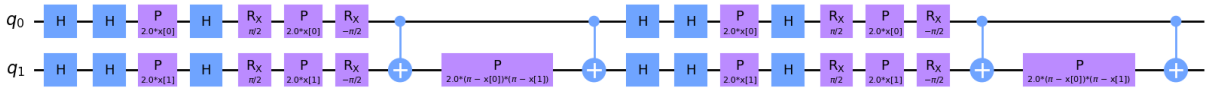

Figure 13: Quantum circuit representation of the PauliFeatureMap for the ad_hoc dataset, depicting the sequential Hadamard and phase gates for initial state preparation, followed by Rx rotations for x-axis encoding, and controlled-phase entanglements, repeated over two iterations to intricately encode classical data into the quantum feature space.

Custom Feature Map Quantum Circuit The CustomFeatureMap for the ad_hoc dataset is constructed with 2 qubits to mirror the binary classification of the data, each representing one class. In this specific configuration, the circuit undergoes four repetitions to thoroughly incorporate the classical data features into the quantum state. The circuit initiates with a sequence of U1 and U2 gates, which are single-qubit rotations that map the classical features onto the quantum state through specific angles in the Bloch sphere, providing a customized approach to encoding.

The CNOT gates between the qubits introduce entanglement, establishing a quantum correlation that is influenced by the feature values from the classical data. The alternating sequence of U1 and U2 gates, followed by CNOT gates, repeats throughout the circuit, each cycle embedding the data more deeply into the quantum state.

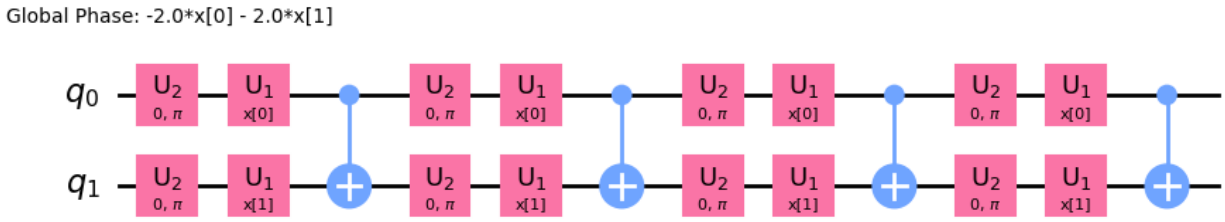

Figure 14: Quantum circuit representation of the CustomFeatureMap for the ad_hoc dataset, highlighting the tailored sequence of U1 and U2 single-qubit rotations for feature encoding, punctuated by CNOT gates for entanglement, iterated across four cycles to deeply encode the classical data into the quantum domain.

4.3.2 Results

This section presents the results of the spectral clustering experiments conducted on the ad_hoc dataset. The performance of each method is quantitatively evaluated using a well-established metric, with the aim of discerning the efficacy of quantum-enhanced spectral clustering against classical approaches. A detailed analysis of the outcomes is provided, offering insights into the capabilities of the proposed quantum clustering methods.

Performance Metric The effectiveness of the clustering algorithms is measured using the Normalized Mutual Information (NMI) score. NMI is an information-theoretic metric that quantifies the quality of clustering by assessing the mutual information between the predicted clusters and the true labels, normalized to account for the size of the dataset. An NMI score of 1 indicates perfect correlation between the clustering outcome and the true labels, while a score of 0 denotes no mutual information. It is a robust metric often used in clustering analysis because it is independent of the absolute values of the labels and is symmetric, ensuring the mutual information is always measured in the same way irrespective of the order of comparison.

Quantitative Results Analyzing the results from the ad_hoc dataset, we see a clear disparity in the performance of the different spectral clustering methods. The Quantum Spectral Clustering with ZZFeatureMap significantly outperforms the other methods, achieving a Normalized Mutual Information Score (NMI) of approximately 0.821. This indicates a strong agreement between the clusters formed by the quantum algorithm and the true labels, highlighting the potential of quantum computing to enhance the feature space for clustering tasks.

In contrast, the other quantum feature maps, Pauli and Custom, yield much lower NMI scores of 0.186 and 0.059, respectively. These results suggest that the specific choice of feature map has a profound impact on the clustering outcome, with ZZFeatureMap being particularly well-suited for this dataset.

The classical spectral clustering methods, both vanilla and PCA-enhanced, perform similarly to each other but are significantly outpaced by the ZZFeatureMap, with an NMI score of just over 0.045. This reinforces the notion that classical methods may struggle with the complex geometries and distributions inherent in the ad_hoc dataset, which are better captured by the quantum approach.

The results are resumed in the following Table:

Method	NMI Score
Quantum Spectral Clustering with ZZFeatureMap	**0.82143**
Quantum Spectral Clustering with PauliFeatureMap	0.18601
Quantum Spectral Clustering with CustomFeatureMap	0.05961
Classic Spectral Clustering Vanilla	0.04528
Classic Spectral Clustering with PCA	0.04528

Table 1: Normalized Mutual Information Scores for different spectral clustering methods on the ad_hoc dataset. The highest score, indicative of the best clustering performance, is achieved by Quantum Spectral Clustering with ZZFeatureMap.

The quantitative results indicate a significant variance in the performance of the spectral clustering methods employed. The Quantum Spectral Clustering with ZZFeatureMap demonstrated a superior ability to discern the complex patterns within the ad_hoc dataset, as reflected in its high NMI score. In contrast, the other quantum methods, and particularly the classical spectral clustering approaches, did not achieve the same level of accuracy.

Qualitative Results The qualitative results provide a visual understanding of the clustering performance. By comparing the predicted labels with the true labels, we can visually assess the clustering algorithm's ability to capture the intrinsic structure of the dataset.

The visualization of the clustering results using the **ZZFeatureMap** shows a high degree of alignment between the predicted and true labels. The plots indicate that the quantum clustering method has effectively identified the underlying groupings within the data. The true labels and predicted labels are nearly indistinguishable, which suggests that the quantum-enhanced method has successfully uncovered the latent structures that classical methods struggled to capture.

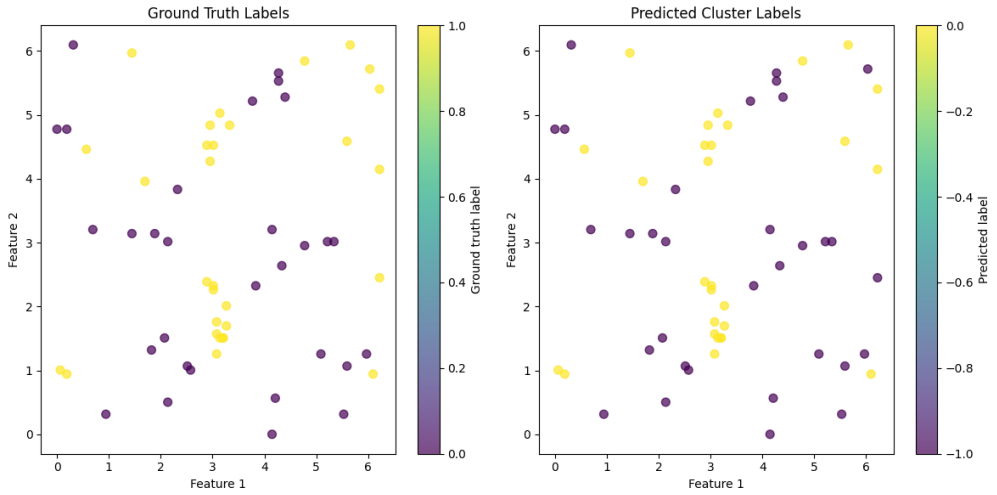

Figure 15: Comparison of true labels (left) and predicted labels (right) using ZZFeatureMap clustering.

Conversely, the **PCA-classical spectral clustering's** visual results present a stark contrast. There is a noticeable discrepancy between the predicted labels and the true labels, highlighting the method's limitations in dealing with complex data structures. The predicted labels plot demonstrates a scattered and unstructured arrangement, lacking the clear definition and separation evident in the true labels plot. This reinforces the quantitative findings and illustrates the challenges classical spectral clustering faces with complex datasets.

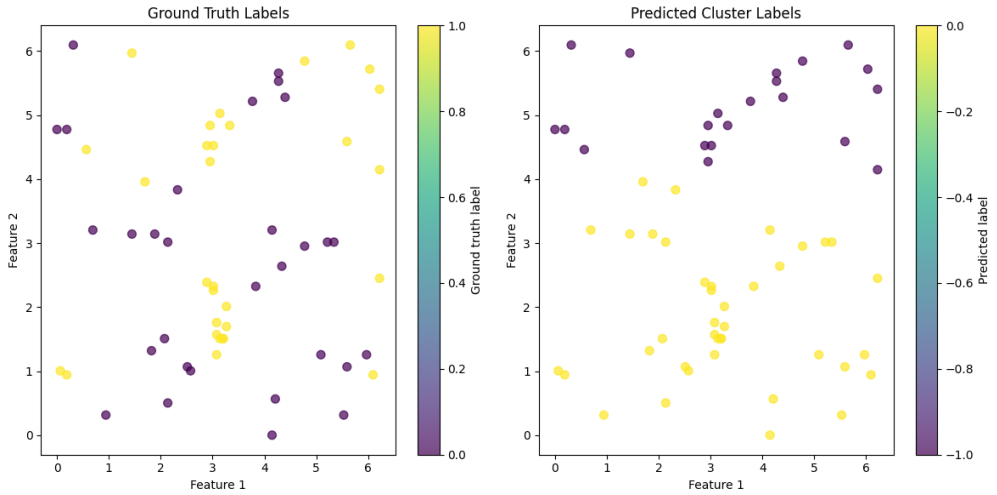

Figure 16: Comparison of true labels (left) and predicted labels (right) using classical spectral clustering.

The side-by-side comparison of the true and predicted labels for both the quantum and classical methods underlines the quantum approach's superior performance. The coherence in the quantum-based visualization is indicative of a more accurate and insightful clustering that can be leveraged for complex datasets.

4.4 Experiments on Synthetic Dataset

Building upon the methodologies and insights gained from the ad_hoc dataset, this section focuses on the synthetic dataset, which presents its unique challenges with intricate three-dimensional structures. The synthetic dataset experiments are critical for assessing the robustness of quantum spectral clustering, especially given the dataset's complex and non-linear nature. As with the ad_hoc dataset, the experiments are delineated based on the dataset to highlight the dataset-specific quantum circuitry employed in the analysis. This section will unveil the quantum circuits tailored for the synthetic dataset, followed by a detailed discussion of the quantitative and qualitative results obtained from these experiments.

4.4.1 Quantum Architecture Description

The synthetic dataset's inherent complexity necessitates a tailored quantum circuit architecture that can capture its non-linear and high-dimensional features. In this section, we delve into the specific quantum circuits utilized for the spectral clustering experiments on the synthetic dataset. Each quantum feature map is carefully constructed with a particular choice of hyper-parameters, such as the number of qubits and the number of repetitions (reps), or in other words, the depth of the circuit. The subsections that follow will outline the architectural details of the ZZFeatureMap, PauliFeatureMap, and CustomFeatureMap, employed in our quantum machine learning pipeline. Accompanying each description, figures of the quantum circuits will be presented to offer a visual comprehension of the discussed architectures.

ZZ Feature Map Quantum Circuit The quantum circuit for the ZZFeatureMap tailored to the synthetic dataset is constructed with 3 qubits to accommodate the three-dimensional nature of the data, with four repetitions to enrich the feature encoding process. Initially, each qubit undergoes a Hadamard gate (H), placing them into a superposition state that allows for the classical data to be mapped onto the quantum system.

Subsequently, parameterized phase rotation gates (P) tailored to the data features are applied to each qubit, thereby embedding the classical information into the quantum domain. The series of controlled-Z (CZ) gates enable the entanglement of the qubits, a key aspect in capturing the inter-feature correlations and complexities present in the synthetic dataset.

The circuit includes multiple layers of these operations, with phase rotations interlaced with entangling CZ gates to ensure a thorough encoding of data into the quantum state. This multi-layer approach is designed to exploit the high-dimensional space that quantum systems offer, providing a rich, intricate representation of the dataset which could potentially reveal patterns and clusters not easily discernible by classical methods.

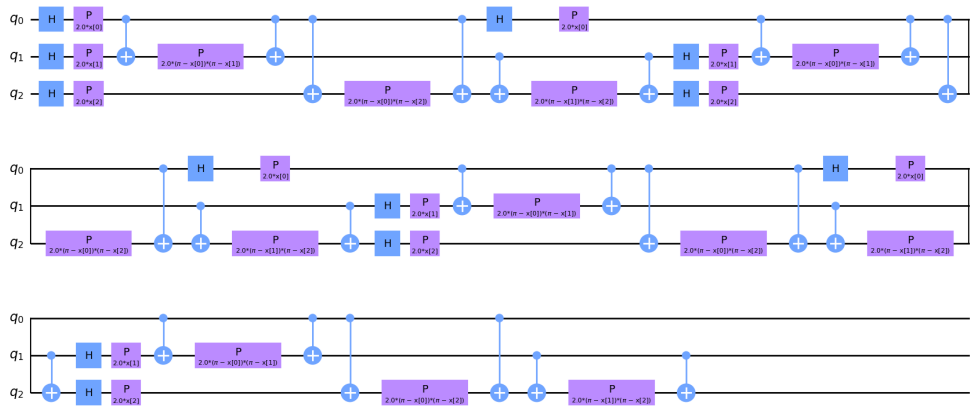

Figure 17: Quantum circuit depiction of the ZZFeatureMap for the synthetic dataset, showing the initial Hadamard gates for state preparation, subsequent parameterized phase rotations, and controlled-Z gates to entangle qubits across four cycles, illustrating the embedding of three-dimensional classical data into a quantum feature space.

Pauli Feature Map Quantum Circuit For the PauliFeatureMap configured for the synthetic dataset, the circuit is constructed using 3 qubits to match the dimensions of the dataset, with each qubit representing a feature dimension. The circuit is initiated with Hadamard gates (H) to create a superposition of states, followed by a sequence of Pauli rotations (P and Rx gates) to encode the classical data points into the quantum state. The circuit is repeated 4 times.

The use of Rx gates, in combination with phase rotation gates (P), provides a rich and complex encoding that is sensitive to the nuances of the synthetic dataset's structure. The entanglement is introduced through controlled operations that link the qubits, allowing the quantum model to capture relationships between data features that might be beyond the reach of classical computing methods.

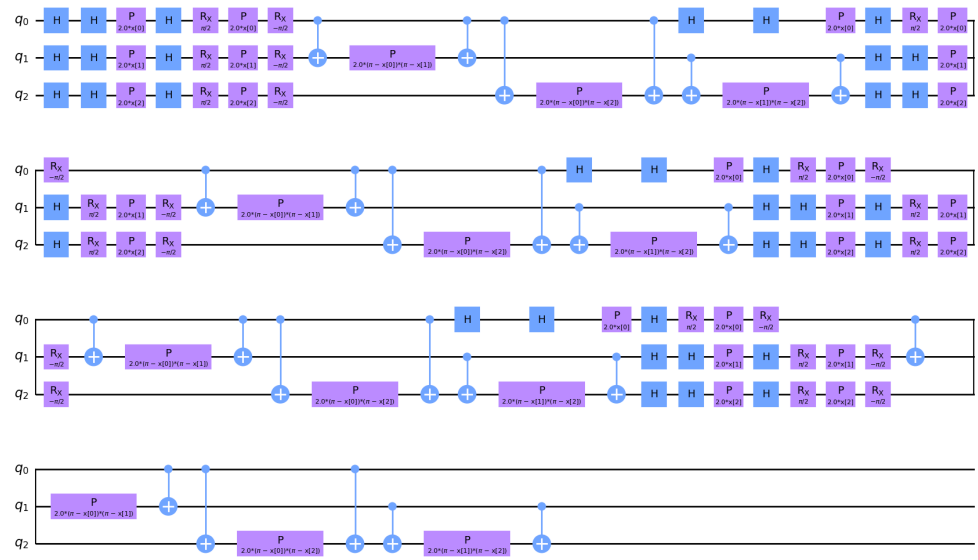

Figure 18: Quantum circuit representation of the PauliFeatureMap for the synthetic dataset, illustrating the complex encoding process through Hadamard and Pauli rotation gates, interspersed with entangling operations to create a high-dimensional quantum feature space.

Custom Feature Map Quantum Circuit For the Custom Feature Map designed for the synthetic dataset, I employ a quantum circuit comprising 3 qubits, reflecting the dataset's three-dimensional nature. The circuit starts with universal single-qubit gates (U1, U2), allowing precise encoding of classical data into the quantum state by adjusting qubit amplitude, phase, and spin. Similarly, this process is repeated 4 times for consistency.

Entanglement among qubits is achieved through a series of controlled operations (CNOT gates), crucial for capturing intricate relationships within data features. This entanglement between qubits is a unique characteristic of quantum systems, enabling the exploration of correlations beyond the reach of classical computational models.

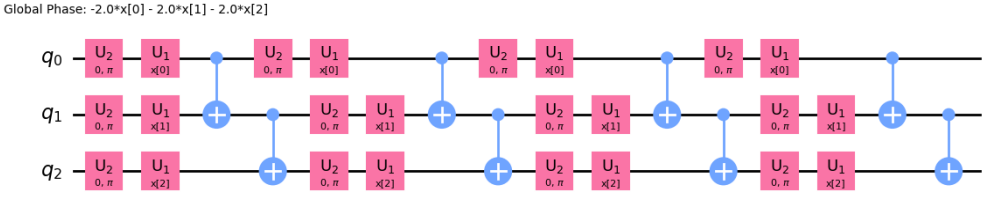

Figure 19: Quantum circuit of the Custom Feature Map designed for the synthetic dataset, depicting the strategic use of universal single-qubit gates and controlled operations to entangle qubits, thereby embedding the classical data into a sophisticated quantum state suitable for clustering analysis.

4.4.2 Results

This section presents the results of the spectral clustering experiments conducted on the *synthetic dataset*. The performance of each method is quantitatively evaluated using a well-established metric, with the aim of discerning the efficacy of quantum-enhanced spectral clustering against classical approaches. A detailed analysis of the outcomes is provided, offering insights into the capabilities of the proposed quantum clustering methods.

Performance Metric As for the experiments conducted on the ad_hoc dataset, the performance metric used it the same, Normalized Mutual Information (NMI) score.

Quantitative Results The results for the synthetic dataset indicate a close competition between the quantum spectral clustering approaches, with the ZZFeatureMap slightly leading the performance, as indicated by the normalized mutual information score. The CustomFeatureMap follows closely behind, demonstrating the effectiveness of a tailored approach. This suggests that the entanglement and superposition offered by the quantum feature maps are effectively utilized to discern the underlying patterns within the data that are less apparent to classical methods.

The performance of the CustomFeatureMap, with an NMI score of 0.6200, is noteworthy as it underscores the potential advantages of customized quantum circuits that are fine-tuned to the specific characteristics of the dataset. It suggests that with further optimization, such custom maps could potentially outperform standard quantum feature maps.

Both quantum methods outperform the classical spectral clustering algorithms, suggesting that quantum-enhanced feature spaces have the potential to better capture the intricate structures of complex synthetic data. The classical methods, while less effective, still offer valuable clustering information, which could be integrated into hybrid approaches for improved performance.

The results of the experiments are shown in the following Table:

Method	**NMI Score**
Quantum Spectral Clustering with ZZFeatureMap	**0.6244**
Quantum Spectral Clustering with PauliFeatureMap	0.5813
Quantum Spectral Clustering with CustomFeatureMap	0.6200
Classic Spectral Clustering Vanilla	0.3899
Classic Spectral Clustering with PCA	0.4388

Table 2: Normalized Mutual Information scores for spectral clustering on the synthetic dataset. The highest score, representing the best clustering performance, is achieved by the Quantum Spectral Clustering with ZZFeatureMap. The second best is achieved by the Quantum Spectral Clustering with CustomFeatureMap.

This comparison highlights a critical aspect of quantum spectral clustering: the quantum advantage seems to manifest more robustly in scenarios where the data's complexity is beyond the reach of classical methods. The results offer a quantitative affirmation that quantum feature maps can provide a richer feature space leading to enhanced clustering results. However, it is also important to note the computational cost associated with quantum methods, which may impact their scalability and practicality for larger datasets.

In summary, these quantitative results suggest that quantum spectral clustering, particularly using feature maps like ZZFeatureMap and CustomFeatureMap, can offer superior performance for complex datasets. However, careful consideration must be given to the trade-offs between computational cost and clustering quality, especially as we progress toward larger-scale quantum computing capabilities.

4.5 Multi-Run Experiments

The multi-run experiments aim to assess the scalability of the best-performing quantum method, **ZZFeatureMap**, against the best classical method, **PCA**, as applied to the ad_hoc dataset. The objective is to understand how these methods perform in terms of accuracy, measured by the Normalized Mutual Information (NMI) score, and computational efficiency, as reflected in execution time. These experiments are critical in evaluating the trade-offs between quantum and classical approaches when the dataset size varies.

Such an analysis is very import for highlighting the practical implications of adopting a quantum computing techniques in machine learning, particularly in scenarios where the complexity and size of the dataset pose significant challenges to classical algorithms. Through a series of planned runs, each encompassing a different dataset size - 10, 44 and 200 points data points - this section aims to investigate the performance dynamics of quantum and classical spectral clustering as they scale. The final goal is to understand whether the quantum method, embodied by the ZZFeatureMap, not only maintains its comparative advantage in accuracy over classical methods, but also does so in a manner that is computationally viable: the goal is to understand the trade off between quantum and classical approach.

4.5.1 Quantitative Results

The multi-run experiments on the ad_hoc dataset reveal significant insights into the scalability and efficiency of the ZZFeatureMap quantum spectral clustering compared to PCA classical spectral clustering. As the dataset size increases from 10 to 200 points, the quantum method consistently outperforms the classical method in terms of Normalized Mutual Information (NMI) score, achieving perfect scores for larger datasets. This suggests that the quantum approach is better suited for capturing the complex relationships and structures within the data, even as the dataset size grows.

However, the computational cost, measured in execution time, also scales significantly for the quantum approach. While for the smallest dataset (10 points), the time difference is marginal, for the largest dataset (200 points), the quantum method takes substantially longer than the classical method. This highlights a trade-off between accuracy and computational efficiency in quantum computing approaches to machine learning.

The results indicate that while quantum spectral clustering provides superior accuracy, especially for larger and more complex datasets, its computational demands necessitate consideration of resource availability and processing time in practical applications.

The results are summarized in the following Table:

Training Size	Quantum Score	Classic Score	Quantum Time (s)	Classic Time (s)
10	0.761	0.0073	0.8656	0.0931
44	**1.0**	0.03996	12.8306	4.4915
200	**1.0**	0.00267	278.4636	11.0491

Table 3: Multi-run experiments results on ad_hoc dataset comparing ZZFeatureMap Quantum Spectral Clustering and PCA Classical Spectral Clustering. Bold scores highlight the best NMI scores achieved.

Why execution time increases with the quantum approach The notable increase in execution time for the quantum spectral clustering method, especially as the dataset size grows, can be attributed to the intrinsic characteristics of simulating quantum circuits on classical hardware. When quantum algorithms, such as those using the ZZFeatureMap for spectral clustering, are executed on a simulator instead of an actual quantum computer, the computational resources required increase significantly with the complexity and size of the quantum circuits. This increase is due to several factors:

- **State Vector Simulation:** Quantum simulators on classical computers work by maintaining and manipulating a state vector that represents the quantum state of the system. As the number of qubits (and thereby the dataset size) increases, the size of this state vector grows exponentially, requiring exponentially more memory and computational power to simulate the quantum operations.

- **Gate Operations**:Each quantum gate operation within the circuit, including those used for feature mapping and entanglement, needs to be simulated mathematically on a classical computer. The complexity of these operations, particularly for entangling gates that introduce correlations between qubits, compounds with larger circuits, leading to longer simulation times.
- **Quantum Measurement:** The process of measuring quantum states, essential for extracting useful information from the quantum circuit, involves simulating the collapse of the quantum state vector according to probabilistic outcomes. For larger circuits, this step becomes more computationally intensive.
- **Circuit Depth:** The repetition (reps) of feature maps, aimed at enhancing the encoding of classical data into quantum states, increases the depth of the quantum circuit. A deeper circuit with more layers of gates necessitates more computational steps in the simulation, further extending the execution time.

While quantum computers inherently leverage quantum parallelism to perform operations on all possible states simultaneously, classical simulators must compute these operations sequentially, leading to a drastic increase in simulation time for larger datasets and more complex circuits.

4.5.2 Qualitative Results

Score Comparison The graphical depiction of clustering performance versus training size demonstrates the effectiveness of the quantum approach. The ZZFeatureMap quantum spectral clustering consistently achieves high clustering scores, showing its strong feature mapping and entanglement abilities. In contrast, the classical PCA-based method doesn't show improvement as training size increases, on this particular dataset.

Figure 20: Comparison of clustering scores between Quantum Spectral Clustering with ZZFeatureMap and Classical Spectral Clustering with PCA on the ad_hoc dataset. The quantum approach demonstrates superior performance that scales effectively with training size.

This qualitative analysis suggests that the quantum method's high-dimensional feature space and the ability to capture complex data correlations offer a distinct advantage over classical methods that rely on linear dimensionality reduction techniques like PCA. Moreover, the quantum approach's ability to achieve a perfect clustering score indicates that it can effectively discern the underlying patterns within the ad_hoc dataset, which are likely non-linear and complex.

Execution Time Comparison The execution time comparison between Quantum Spectral Clustering with ZZFeatureMap and Classical Spectral Clustering with PCA reveals a notable difference in computational efficiency. The classical method maintains a relatively consistent execution time, demonstrating only a moderate increase as the training size grows. This is characteristic of classical algorithms, where computational complexity scales more predictably with the size of the dataset.

In contrast, the quantum approach shows a more significant increase in execution time as the training size escalates. While the quantum method's execution time starts lower than the classical method for the smallest training size, it surpasses the classical time at a training size of 44 data points and continues to rise sharply. This exponential increase reflects the higher computational demands of simulating quantum circuits, especially as the feature space expands with more data points.

This qualitative result underscores the current limitations of quantum computational resources, particularly when using simulators that do not benefit from quantum speedup. It also highlights the importance of quantum hardware advancements and algorithmic optimizations to improve the scalability of quantum machine learning methodology.

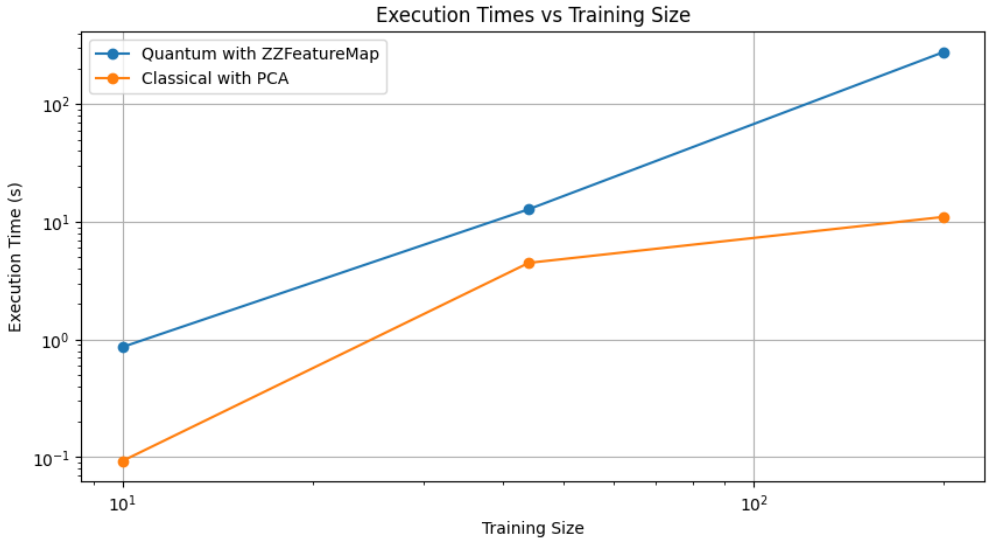

Figure 21: Execution time comparison between Quantum Spectral Clustering with ZZFeatureMap and Classical Spectral Clustering with PCA on the ad_hoc dataset. The graph illustrates the quantum method's rising computational demands with increased training size.

5 Conclusion

In this study, I investigated the potential advantages of integrating quantum computing techniques with spectral clustering algorithms. For this project, an innovative hybrid approach was proposed and evaluated, leveraging quantum kernels and quantum feature maps, while maintaining the classical spectral clustering algorithm structure. The methodology was assessed through extensive experimentation on two challenging datasets: the *ad_hoc* dataset from Qiskit and a *synthetic* complex dataset, generated for this particular study. Moreover, three quantum feature map approaches were evaluated, paired with Fidelity quantum kernel, and for classical methods, both vanilla spectral clustering and a PCA-enhanced version were studied. The objective of the experimentation was to determine whether quantum-enhanced machine learning could outperform classical methods in terms of accuracy and also to examine the trade-offs in computational resources.

The exploration revealed several key findings. First of all, quantum spectral clustering, particularly when utilizing the ZZFeatureMap, demonstrated a superior ability to identify clusters within the *ad_hoc* dataset, as evidenced by the normalized mutual information score. This success was replicated on the synthetic dataset, indicating a consistent advantage of the quantum approach over classical methods for datasets characterized by non-linear separability and high complexity.

Furthermore, the study highlighted the computational trade-offs associated with quantum and classical methods. While quantum spectral clustering achieved higher clustering accuracy on challenging datasets, the computational time required increased exponentially with the number of data points. This behavior underlines the current limitations of quantum simulation on classical hardware and the necessity for quantum hardware development to realize the theoretical speedup of quantum algorithms.

In conclusion, the findings from this study contribute to the growing body of knowledge in quantum machine learning. They affirm the potential of quantum spectral clustering to surpass classical techniques in certain situations, particularly in dealing with data that exhibit complex, non-linear structures. Nonetheless, the practicality of quantum methods, especially in terms of scalability and computational efficiency, remains a pressing challenges.

As we look forward, the next steps in research should focus on enhancing the scalability of quantum clustering methods, exploring adaptive feature maps that can be tailored to specific data structures, and developing error mitigation strategies to improve the reliability of quantum computations. Further exploration into hybrid quantum-classical algorithms may also yield improvements, allowing for more efficient utilization of quantum resources.

References

Vojtěch Havlíček, Antonio D Córcoles, Kristan Temme, Aram W Harrow, Abhinav Kandala, Jerry M Chow, and Jay M Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747): 209–212, 2019.

Qiskit contributors. Qiskit: An open-source framework for quantum computing, 2023.

Andrew Ng, Michael Jordan, and Yair Weiss. On spectral clustering: Analysis and an algorithm. *Advances in neural information processing systems*, 14, 2001.

Seth Lloyd, Maria Schuld, Aroosa Ijaz, Josh Izaac, and Nathan Killoran. Quantum embeddings for machine learning. *arXiv preprint arXiv:2001.03622*, 2020.