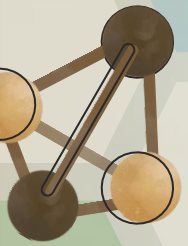
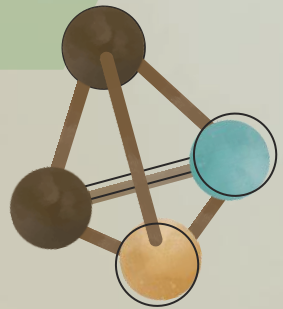
FUNDACJA QUANTUM AI

Q Poland

Quantum Hackathon Challenge: Kernel Methods for Molecular Classification





- Rihab Hoceini
- Raouf Ould Ali
- Amieur Lilya Fatima-Zohra
- Widad Hassina Belkadi
- Hamza Abderaouf Khentache

QFennecs

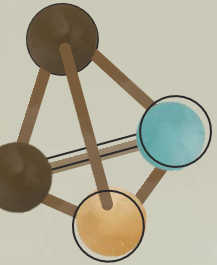
Introduction

Graph Structured Data

Many real-world data are best represented as **graphs**, capturing relationships between entities. making it ideal for **molecular**, biological, and plant data analysis.

The need of feature Map

In real-world domains , relying solely on raw graph structures is insufficient for effective learning Feature maps are needed to transform graphs into informative, high-dimensional representations that highlight key properties



What missing in Current methods

Limited Transferability Across Molecules

Feature maps optimized for one dataset or molecular class often fail to generalize to new compounds or biological systems.

Missing Domain Properties

Current mappings mainly rely on topological and mathematical structure without integrating physical or chemical parameters such as electron density, energy levels, or interaction forces.

Redundancy and Scalability Issues

Feature maps for graphs frequently generate repetitive structural information across nodes and edges, creating highly correlated features that add little new insight.

Overall Pipeline

Data Loading

Graph conservation

Cross Validation

Feature extraction

SVM based Kernel

Data Loading → Graph conservation → Cross Validation → Feature extraction → SVM based Kernel

Dataset	Node Features	Edge Features	Target (Y)	Graph Count	Nbr of unique nodes
PTC-MR	Atom type	Bond type	Toxic (1) / Non-toxic (0)	~344	19
NCI1	Atom type	Bond type	Active (1) / Inactive (0)	4110	37
MUTAG	Atom type	Bond type	Mutagenic (1) / Non-mutagenic (0)	188	7
AIDS	Atom type	Bond type	Active (1) / Inactive (0)	2000	38
PROTEINS	Node type continuous attributes	Edge type	Enzyme (1) / Non-enzyme (0)	1113	3

$$G = (V_i, E_i)$$

Data Loading → **Graph conservation** → Cross Validation → Feature extraction → SVM based Kernel

Each molecule is represented as an undirected graph

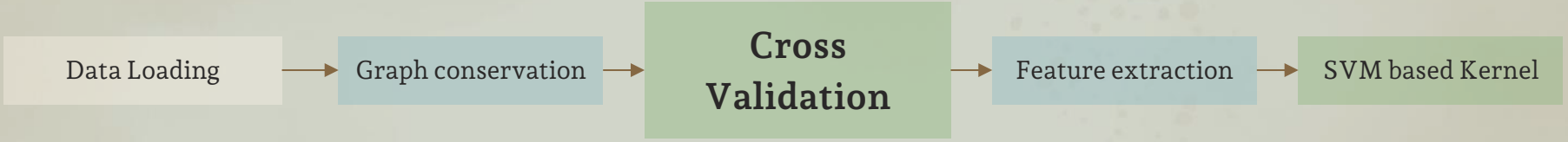
$$G = (V, E)$$

the set of atoms (nodes)

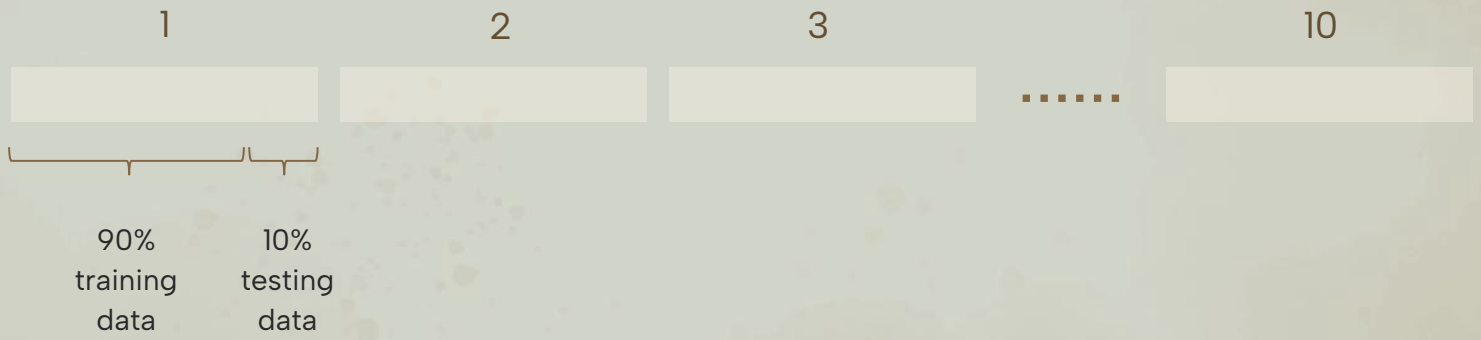
the set of chemical bonds (edges)

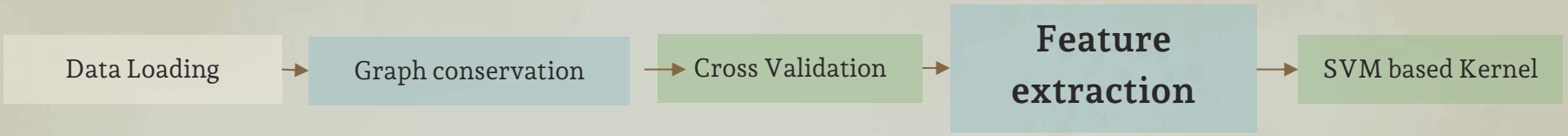
$v_i \in V$

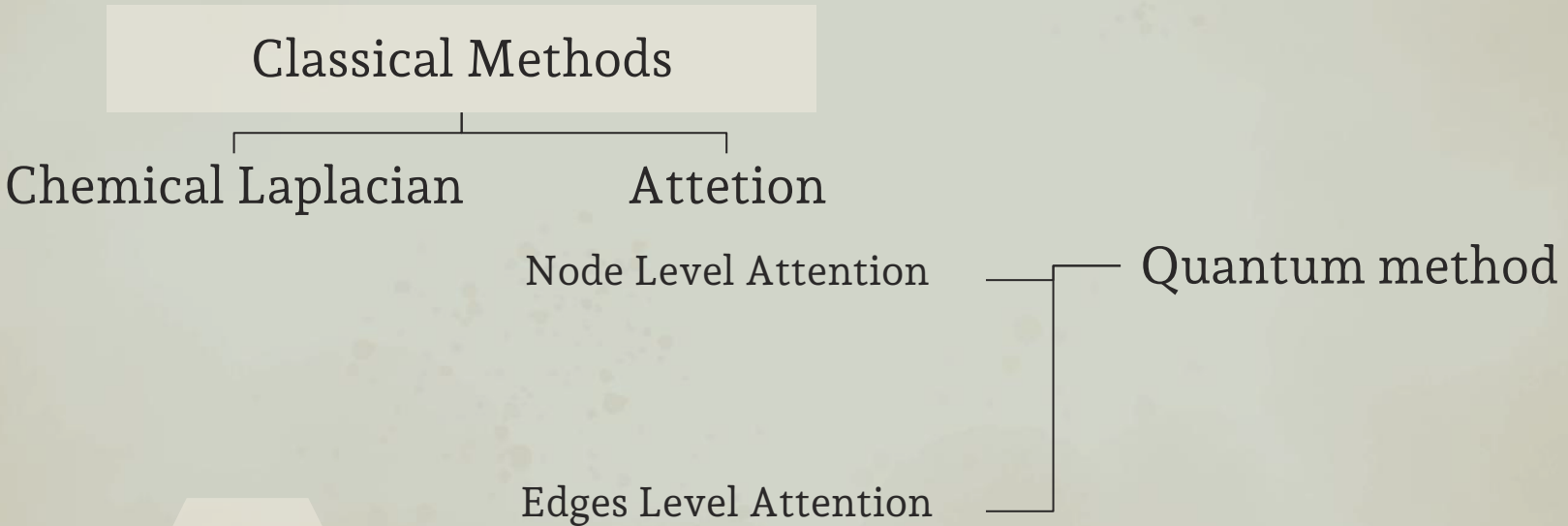
$(v_i, v_j) \in E$



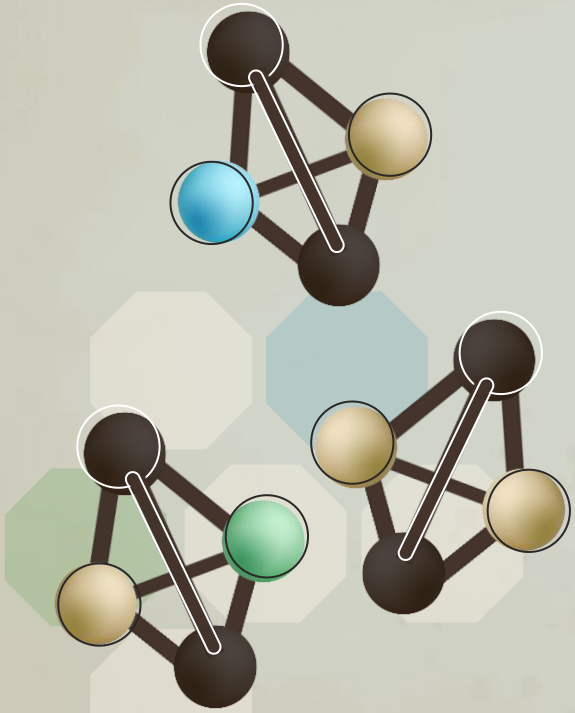
10 folds using stratified cross-validation

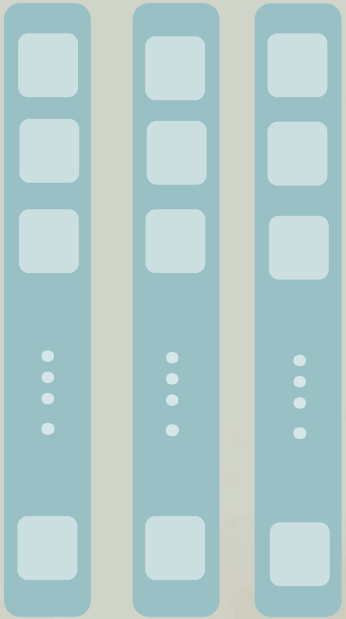



Data Loading
Graph conservation
Cross Validation
Feature extraction
SVM based Kernel


Classical Methods
Chemical Laplacian
Attetion
Node Level Attention
Edges Level Attention
Quantum method

Feature Map





Classification

Chemical Laplacian

$$L_{chem} = \begin{cases} \alpha(v_i) \cdot \sum_k w(i,k), & i = j \\ \gamma\big(w(i,j)\big) \cdot w(i,j), & (i,j) \in E \\ 0, & otherwise \end{cases}$$

→ Weighted degree of atom i

→ bond strength and bond order

⇩

Compute its eignvalues

$$L_{chem}x = \lambda\, x$$

⇩

Final Output

$[\lambda'_1, \lambda'_2 \ldots . \lambda'_{10}, \text{trace\_norm}, \text{algebraic\_conn}]$

Attention Based Feature Map

Node-level Attention

For each atom $v_i \in V$,
we build a feature vector x_i

$$x_i = [t_i, \deg(i), C_i, N_i]$$

⇩

$$a_i = Softmax\left(\frac{(x_i W_q \cdot x_j W_k)^T}{\sqrt{d}}\right)$$

quantify the relative importance of each atom in its local neighborhood

Edge-level Attention

For each bond between atoms (u, v)
we form an edge feature vector e_{uv}

$$e_{uv} = [b_{type}, t_u, t_v, \deg(u), \deg(v)]$$

⇩

$$a_{uv} = \sigma(e_{uv} W_e W_e^{\intercal} e_{uv}^{\intercal})$$

computes a **self–similarity score** (a measure of how important this bond is)

Quantum Based

Graph-to-Quantum Encoding

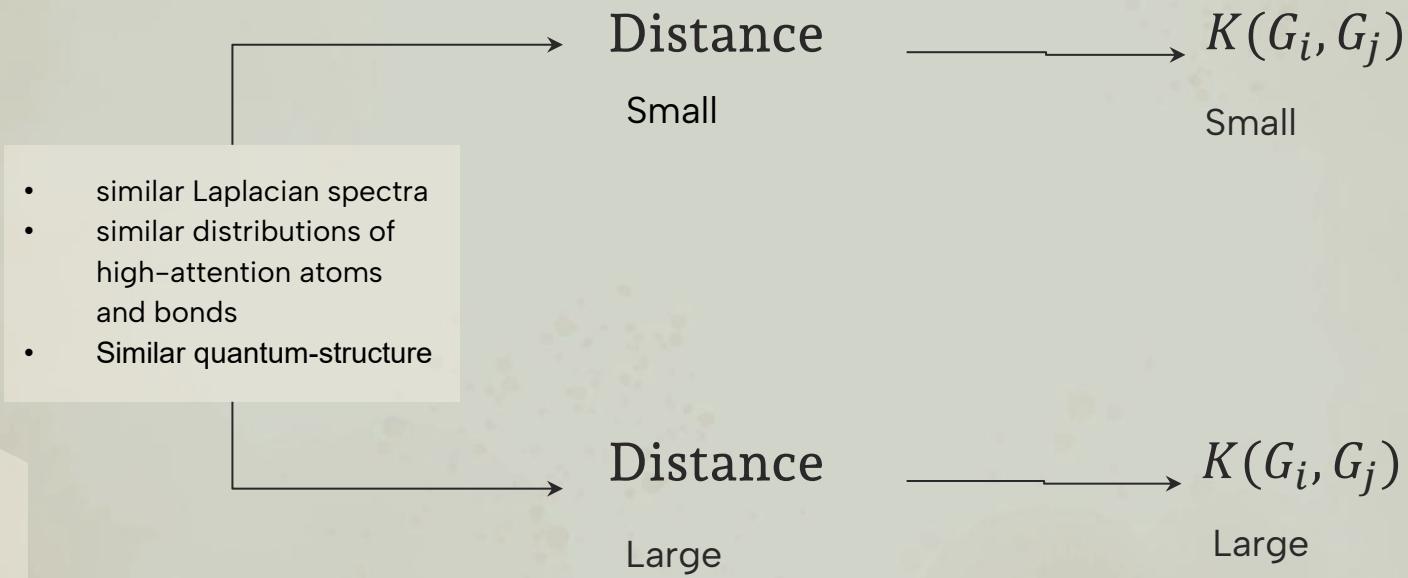
Each molecular graph G = (V, E) is transformed into a quantum graph state

$$|G\rangle = \prod_{(i,j)\in E} CZ_{ij} \prod_{v_i \in V} H_i |0\rangle^{\otimes |V|}$$

Quantum observable

- Entanglement Entropy
- Quantum Walk Overlap
- Molecular Hamiltonian Expectation

How our methods informative for classification




Data Loading
Graph conservation
Cross Validation
Feature extraction
SVM based Kernel
normalization using the scaling transformation
train a support vector machine (SVM) with an **RBF kernel**

Results

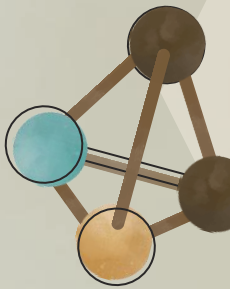
Table 2: Comparison of F1-score and cross-validation (CV) accuracy across feature extraction methods and datasets.

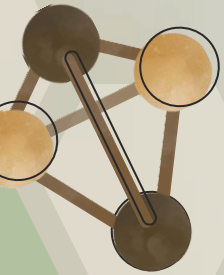
Dataset	Chemical Laplacian		Attention-Based		Quantum + Attention-Based	
	F1	CV Acc.	F1	CV Acc.	F1	CV Acc.
MUTAG	0.8	0.82	0.83	0.85	**0.90**	**0.86**
PTC-MR	0.52	0.59	0.51	0.58	**0.53**	**0.60**
NCI1			0.70	0.73	**0.73**	**0.74**
AIDS	0.84	0.95	0.98	0.98	**0.99**	**0.98**
PROTEINS			0.63	0.66	**0.60**	**0.67**

Is **Quantum** **THE ONE** for structured graph data?

Conclusion

Our study we built 3 feature map methods : Chemical Laplacian, Attention, and Quantum feature maps for molecular classification.
The Quantum + Attention model achieved the best performance, showing that merging structural, chemical, and quantum information creates richer, more discriminative molecular representations





Thank *you*