

Machine learning mutation-acyclicity of quivers

Kymani T.K. Armstrong-Williams^{a,*}, Edward Hirst^a, Blake Jackson^b, Kyu-Hwan Lee^{b,c}

^a Centre for Theoretical Physics, Queen Mary University of London, E1 4NS, UK

^b Department of Mathematics, University of Connecticut, Storrs, CT 06269, USA

^c Korea Institute for Advanced Study, Seoul 02455, Republic of Korea

ARTICLE INFO

Dataset link: https://github.com/KTKAW/MACHINE_LEARNING_MUTATION_ACYCLICITY_OF_QUIVERS.git

Keywords:

Cluster algebras
Quiver mutation
Mutation-acyclicity
Machine learning
Computer-assisted proof

ABSTRACT

Machine learning (ML) has emerged as a powerful tool in mathematical research in recent years. This paper applies ML techniques to the study of quivers—a type of directed multigraph with significant relevance in algebra, combinatorics, computer science, and mathematical physics. Specifically, we focus on the challenging problem of determining the mutation-acyclicity of a quiver on 4 vertices, a property that is pivotal since mutation-acyclicity is often a necessary condition for theorems involving path algebras and cluster algebras. Although this classification is known for quivers with at most 3 vertices, little is known about quivers on more than 3 vertices. We give a computer-assisted proof of a theorem to prove that mutation-acyclicity is decidable for quivers on 4 vertices with edge weight at most 2. By leveraging neural networks (NNs) and support vector machines (SVMs), we then accurately classify more general 4-vertex quivers as mutation-acyclic or non-mutation-acyclic. Our results demonstrate that ML models can efficiently detect mutation-acyclicity, providing a promising computational approach to this combinatorial problem, from which the trained SVM equation provides a starting point to guide future theoretical development.

1. Introduction

In the current era of artificial intelligence (AI) transforming nearly every aspect of society and science, mathematics research is no exception. The efficiency and capability of machine learning (ML) can undoubtedly be utilized in mathematics research. Indeed, intensive research has been conducted on “ML mathematics” to explore whether ML can detect underlying mathematical structures. Through applying ML tools to various datasets generated from mathematical objects, high accuracies (often exceeding 98%) have been achieved in classifying mathematical objects according to their invariants and properties. Examples include algebraic curves [28,30,32,31], special holonomy manifolds [2,3,6,12,13,33], branes [5,10,16,17,49], and number fields [4,29]. Interpreting what the machine learns can sometimes lead to new discoveries or perspectives in the study of mathematical objects. The recent discovery of *murmuration* phenomena in arithmetic is a prime example [19,32]. This proposes a new paradigm in mathematics research: generate datasets, apply ML tools, interpret the results, gain new perspectives, formulate conjectures, and prove them.

In this paper, we apply this framework to the problem of quiver mutations. A *quiver* is a directed multigraph without loops or directed

2-cycles, and quivers are ubiquitous in mathematics, mathematical physics, and computer science, as they efficiently encode fundamental skew-symmetric structures. The *mutation* of a quiver is a combinatorial operation that originates from cluster algebras [25], where it is an essential component, and from quantum field theory, where it relates to Seiberg duality [23,48]. The iterative application of these mutations generates a rich and intricate algebraic combinatorial structure.

When a quiver is obtained from another via successive mutations, we say the two quivers are *mutation-equivalent*. In the theory of cluster algebras and related topics, it is crucial to know whether a quiver is mutation-equivalent to a quiver with no oriented cycles, called an *acyclic* quiver. When a quiver is mutation-equivalent to an acyclic quiver, it is “in control,” and we understand relatively well the constructions based on such a quiver from various tools coming from representation theory and combinatorics. Otherwise, it is “out of control,” and our understanding of the related structures is limited. Thus, knowing whether a quiver is mutation-equivalent to an acyclic quiver is often a question with a highly consequential answer.

However, even for the problem of determining *mutation-acyclicity*—let alone general mutation equivalence—we currently lack practical meth-

* Corresponding author.

E-mail addresses: k.t.k.armstrong-williams@qmul.ac.uk (K.T.K. Armstrong-Williams), e.hirst@qmul.ac.uk (E. Hirst), blake.jackson@uconn.edu (B. Jackson), khlee@math.uconn.edu (K.-H. Lee).

<https://doi.org/10.1016/j.jaca.2025.100040>

Received 11 December 2024; Received in revised form 10 August 2025; Accepted 22 August 2025

ods or algorithms when the number of vertices is greater than 3. With the advent of new approaches from the ML framework, we investigate in this paper whether ML can assist in addressing this problem. After generating datasets from mutations of quivers, we apply ML tools to this problem from both supervised and unsupervised learning. More precisely, we apply neural networks (NNs), support vector machines (SVMs), and principal component analysis (PCA). This extends work from [9,18,20] where NNs were first applied to differentiate mutation classes and their representations, now in this work with a focus on the mutation-acyclicity property.

In the construction of quiver data for analysis, we prove a new Theorem, Theorem 2.9, via computer-assisted means; classifying all rank 4 quivers with edge weight at most 2 according to the mutation-acyclicity property. The results of our following ML experiments clearly show that ML architectures can be used to detect mutation-acyclicity efficiently. We achieve high accuracy in distinguishing mutation-acyclic quivers from non-mutation-acyclic ones using both NNs and SVMs. In fact, it was these first promising ML results that motivated the authors to perform the work required to prove the new Theorem. The SVMs provide directly interpretable equations (viewed as separating hypersurfaces) much like the Markov constant for rank 3 quivers [11]. These trained equations should provide theoretical insight into the underlying structure of high-rank mutation-acyclicity invariants, where we leave full analysis and design of the invariant to future work.

For the ML experiments of this paper, quiver data was generated using the SageMath [50] package [43], and graph analysis was conducted using networkx [27] and graph-tool [46]. The subsequent machine learning tasks were performed in python, utilizing the scikit-learn [45] and tensorflow [1] packages. Scripts were adapted from previous works [18,20], with high-performance computing resources [36]. All scripts and data are made available on this work's respective GitHub.¹

After this introduction, §2 covers background material. In §2.1, quiver mutation and cluster algebras are defined, and some results on mutation-acyclicity are presented, including our new result (Theorem 2.9). In §2.2, the ML methods and tools that are adopted in this paper are reviewed, including NNs, SVMs, and PCA. §3 explains how we generate our datasets from quiver mutations. The next section, §4, contains the results of our ML investigations. §4.1 summarises what we obtain from NNs. §4.2 highlights separating hyperplanes coming from SVMs. In the final section, §5, we make concluding remarks.

2. Background

Cluster algebras provide a well-motivated arena for testing the efficiency of machine learning (ML) methods, chiefly due to the combinatorial nature of the quivers and their matrix representations. A prominent problem within cluster algebras is identifying whether a given quiver is equivalent to an acyclic² quiver or not under the mutation operation. Whilst this is infeasible by eye, this paper examines how effective ML architectures are at learning this property. In this background section, the relevant mathematical details underpinning cluster algebras and this mutation-acyclicity problem are introduced, along with the implemented ML methods.

2.1. Cluster algebras

Cluster algebras are subrings of the field of rational functions in n commuting variables over \mathbb{Q} . They often appear as coordinate rings of certain geometric objects, and were formalized by Fomin and Zelevinsky

¹ https://github.com/KTKAW/MACHINE_LEARNING_MUTATION_ACYCLICITY_OF_QUIVERS.git.

² In graph theoretic terms, we are interested in *strongly* acyclic weighted directed graphs.

[25] in order to study their (dual) canonical bases and total positivity. Since their introduction, cluster algebras have found applications in various areas of mathematics and mathematical physics, including representation theory, combinatorics, algebraic geometry, dynamical systems, knot theory, and string theory.

What makes cluster algebras distinctive is that they are defined recursively via a local process, called *mutation*, with some initial algebraic and combinatorial data, called an *initial seed*. By repeatedly mutating the initial seed (sometimes indefinitely), we can generate the full list of generators required to define a cluster algebra.

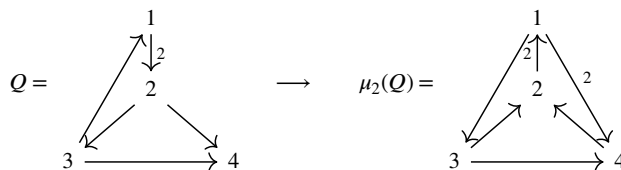
Definition 2.1. A *quiver* is a finite, directed multigraph $Q = (Q_0, Q_1)$. We always assume that a quiver does not have loops or oriented 2-cycles. The elements of $Q_0 = \{1, \dots, n\}$ are the *vertices* of Q and the elements of Q_1 are the *arrows* of Q . In this case, we call n the *rank* of Q . There exist two maps, the *source map* and the *target map* $s, t : Q_1 \rightarrow Q_0$, where $s(\alpha)$ returns the vertex at the source of α and $t(\alpha)$ returns the vertex at the target of the arrow α . We say a quiver is *acyclic* if it contains no directed cycles of any length.

Definition 2.2. Given a quiver Q and a vertex k , we can define a new quiver $\mu_k(Q)$, called the *mutation of Q at vertex k* , in the following way:

- (1) for each directed 2-path $i \rightarrow k \rightarrow j$ in Q , add an arrow $i \rightarrow j$ in $\mu_k(Q)$,
- (2) reverse the direction of all arrows incident to k ,
- (3) pairwise delete any oriented 2-cycles in $\mu_k(Q)$ which have appeared as a result of step 1.

We say a quiver Q' is *mutation-equivalent* to Q if there is a sequence of vertices $[k_1, k_2, \dots, k_\ell]$ such that $Q' = \mu_{k_\ell}(\dots \mu_{k_2}(\mu_{k_1}(Q)) \dots)$: in other words, we can pass from Q to Q' by successive mutations at the vertices k_1, k_2, \dots, k_ℓ . The *mutation class* $[Q]$ of Q is the collection of all quivers Q' that are mutation-equivalent to Q .

Example 2.3. The following is an example of mutation in a rank 4 quiver:



Arrows with the label 2 indicate two arrows between the vertices. As seen from the example and definition, mutating a quiver Q at any vertex k produces another quiver. Moreover, mutation is an involution, meaning that $\mu_k(\mu_k(Q)) = Q$ for any vertex k . It should also be noted that quivers are in bijective correspondence with skew-symmetric integral matrices. If Q is a rank n quiver with $Q_0 = \{1, \dots, n\}$, then Q corresponds to a $n \times n$ skew-symmetric matrix $B = B(Q) = (b_{ij})$ where

$$b_{ij} = (\# \text{ arrows from } i \text{ to } j) - (\# \text{ arrows from } j \text{ to } i).$$

This is the signed adjacency matrix of the quiver when viewed as a directed graph. There is also a notion of mutation of skew-symmetric matrices, and this definition agrees with the definition of mutation of quivers. Since the correspondence between quivers and skew-symmetric integral matrices is bijective and mutation works the same in both cases, we prefer to use the quiver perspective when visualizing quivers and the skew-symmetric matrix perspective for ML.

For Q at the beginning of this example, we have

$$B(Q) = \begin{pmatrix} 0 & 2 & -1 & 0 \\ -2 & 0 & 1 & 1 \\ 1 & -1 & 0 & 1 \\ 0 & -1 & -1 & 0 \end{pmatrix} \quad \text{and}$$

$$\mu_2(B(Q)) = B(\mu_2(Q)) = \begin{pmatrix} 0 & -2 & 1 & 2 \\ 2 & 0 & -1 & -1 \\ -1 & 1 & 0 & 1 \\ -2 & 1 & -1 & 0 \end{pmatrix}$$

Definition 2.4. Let $\mathbf{x} = \{x_1, \dots, x_n\}$ be a set of n variables and Q be a quiver on n vertices, where the vertex labeled k is understood to correspond with the variable x_k . Let $\mathcal{F} = \mathbb{Q}(x_1, \dots, x_n)$ be the field of rational functions in \mathbf{x} . *Mutation of the seed* (\mathbf{x}, Q) at vertex k is the pair $\mu_k(\mathbf{x}, Q) = (\mathbf{x}', Q')$ where $Q' = \mu_k(Q)$ and $\mathbf{x}' := \mu_k(\mathbf{x}) = (\mathbf{x} \setminus \{x_k\}) \cup \{x'_k\}$ and $x'_k \in \mathcal{F}$ is defined by the *exchange relation*:

$$x_k x'_k = \prod_{\alpha: s(\alpha)=k} x_{t(\alpha)} + \prod_{\alpha: t(\alpha)=k} x_{s(\alpha)}.$$

We call $\mathbf{x}'' \in \mathcal{F}$ a *cluster* if $\mathbf{x}'' = \mu_{k_\ell} \dots \mu_{k_2} \mu_{k_1}(\mathbf{x})$ for some sequence $[k_1, k_2, \dots, k_\ell]$ and elements $x''_k \in \mathbf{x}''$ *cluster variables*. The *cluster algebra* (of rank n) $\mathcal{A} = \mathcal{A}(\mathbf{x}, Q)$ is the \mathbb{Z} -subalgebra of \mathcal{F} generated by all the cluster variables.

The role that the initial combinatorial data (\mathbf{x}, Q) plays in the structure of $\mathcal{A}(\mathbf{x}, Q)$ cannot be overstated. The addition or removal of a single arrow of Q can determine whether the resulting algebra consists of finitely many or infinitely many cluster variables, and many theorems in the cluster algebra literature depend on the mutation class of Q . Specifically, there is a collection of results that require the initial quiver Q to be acyclic, so that the cluster algebra is well-behaved.

2.1.1. *Quiver mutation-acyclicity*

In this section, we define mutation-acyclicity and motivate it with examples from the literature. Going through each of these examples in detail is beyond the scope of this paper, and we direct readers to the references in this section should they want to investigate further.

Definition 2.5. Let Q be a quiver. We say that Q is *mutation-acyclic* if there exists at least one acyclic quiver in $[Q]$, or equivalently, if Q is mutation-equivalent to an acyclic quiver Q' . If there are no acyclic quivers in $[Q]$, then we say Q is *non-mutation-acyclic*.

Researchers care about acyclic and mutation-acyclic quivers for a variety of reasons. For one, path algebras over acyclic quivers have finite vector space dimension [8,47]. These are important objects of study since every hereditary Artin algebra is Morita equivalent to a path algebra over some quiver, and finite matrices can represent the maps between finite-dimensional algebras. Buan, Marsh, and Reiten show that, for a cluster algebra arising from an acyclic quiver, cluster mutations correspond to tilting objects in the cluster category over the path algebra of Q [15]. In terms of the cluster algebras this means that, if Q is acyclic, the denominator vectors of the non-initial cluster variables of $\mathcal{A}(\mathbf{x}, Q)$ correspond to the dimension vectors of the indecomposable rigid representations of Q over the path algebra kQ , where k is an algebraically closed field; these dimension vectors are the so-called *real Schur roots*. Furthermore, these denominator vectors (real Schur roots) correspond to the positive roots of the root system of the Kac–Moody algebra associated with Q [14,34]. It was also shown that the real Schur roots coincide with the c -vectors of the cluster algebra $\mathcal{A}(\mathbf{x}, Q)$ when Q is acyclic [44]. There has been research attempting to understand the c -vectors of cluster algebras generated from acyclic quivers for this reason, including some work by the last two authors [21,22,39,40].

Another concept is that of *reddening sequences* in cluster algebras with coefficients, and it is known that mutation-acyclic quivers admit a reddening sequence. In his paper introducing reddening sequences (maximal green sequences), Keller [35] shows that reddening sequences produce explicit formulas for the refined Donaldson–Thomas invariants introduced by Kontsevich and Soibelman [38]. The existence of reddening sequences is also a sufficient condition for the existence of a theta basis for the upper cluster algebra (a sought-after canonical basis) [26].

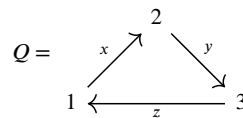
Other research focuses on the class of *locally acyclic cluster algebras* [41] where the cluster algebra can be “covered” by cluster algebras generated by acyclic quivers. These have many desirable properties, chief among them that the cluster algebra is equal to the upper cluster algebra ($\mathcal{A} = \mathcal{U}$), which is the intersection of Laurent rings of the cluster variables and has a geometric interpretation [42].

Given the desirable properties that cluster algebras generated from acyclic quivers possess, one might want to know when a given quiver is mutation-equivalent to an acyclic quiver; in other words, when a quiver is mutation-acyclic. This is because $\mathcal{A}(\mathbf{x}, Q') \cong \mathcal{A}(\mathbf{x}, Q)$ when Q and Q' are mutation-equivalent. However, deciding whether or not a given quiver is mutation-acyclic is a difficult problem; in fact, there is no known algorithm or invariant for detecting mutation-acyclicity in quivers with more than 3 vertices (except in special cases). We now survey the known results that can be used to determine mutation-acyclicity.

2.1.2. *Quiver mutation-acyclicity in rank 1, 2, and 3*

The problem is trivial for rank 1 and rank 2 quivers since they are all, by definition, acyclic, and so every quiver of rank 1 or 2 is mutation-acyclic. The first non-trivial case occurs when Q is a cyclic quiver of rank 3.

In 2008, Assem, Blais, Brüstle, and Samson [7] described an algorithm for detecting mutation-acyclicity³ in rank 3 quivers. Let



be a cyclic quiver on 3 vertices. The algorithm described is simple: mutate at the vertex opposite the largest collection of edges. So, if z were the largest collection of edges in Q , mutate at vertex 2. If there is more than one largest collection of edges, the one you pick to mutate across from doesn't matter. This process continues until either an acyclic quiver is reached or mutation at any vertex does not decrease the total number of edges $(x + y + z)$.

In 2011, Beineke, Brüstle, and Hille [11] were able to circumvent the algorithm described above and produced a mutation-invariant constant (called the Markov constant) which is used to classify rank 3 quivers as mutation-acyclic or not. Let

$$C(x, y, z) = x^2 + y^2 + z^2 - xyz$$

be the *Markov constant* associated to a rank 3 cyclic quiver Q (as above). A partial statement⁴ of their two main theorems is below.

Theorem 2.6. [11, Theorem 1 and 2] *Let Q be a cyclic quiver on 3 vertices with numbers of arrows given by $x, y, z \in \mathbb{Z}_{\geq 0}$. Then Q is mutation-acyclic if any of the following are satisfied:*

- (1) $C(x, y, z) > 4$ or $\min\{x, y, z\} < 2$.

The quiver Q is non-mutation-acyclic if any of the following are satisfied:

- (1) $C(x, y, z) < 0$
- (2) $C(x, y, z) \leq 4$ and $x, y, z \geq 2$.

A consequence of this theorem is that the only non-mutation-acyclic quiver of rank 3 with at most 2 arrows between any pair of vertices is the Markov quiver:

³ This algorithm, along with the fact that the acyclic quivers in a mutation-class form a connected component of the exchange graph, can actually be used to detect mutation-equivalence in rank 3 quivers. This algorithm results in a minimal representative of a quiver's mutation class.

⁴ The full pair of theorems has a few more conditions, but for our purposes, the statement below is sufficient.

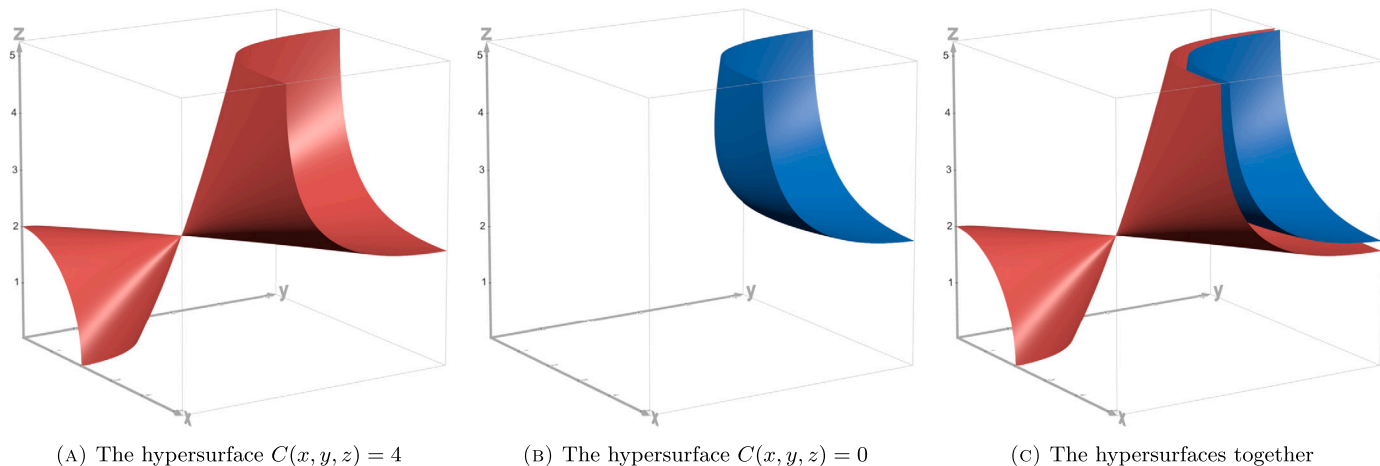
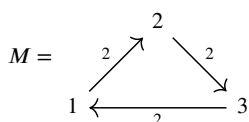


Fig. 2.1. The Markov constant classification visualized as separating hypersurfaces.



The ability to check whether a rank 3 quiver is non-mutation-acyclic using the Markov constant plays a large part in the proof of Theorem 2.9 and in generating data for our ML experiments.

As a foreshadowing of the ML techniques used later in this paper, Theorem 2.6 has a geometric interpretation. There are two “critical values” for the Markov constant: 0 and 4. If $C(x, y, z) > 4$ (or there is a single/missing arrow), then the quiver is mutation-acyclic; and if $C(x, y, z) < 0$, then the quiver is non-mutation-acyclic. By plotting the triple (x, y, z) of edge weights in the first octant, we can see that $C(x, y, z) = 0, 4$ defines two separating (or classifying) nonlinear hypersurfaces in \mathbb{R}^3 (Fig. 2.1). The singularity in Fig. 2.1a occurs at the point $(2, 2, 2)$, which is the Markov quiver. This geometric classification was one of the primary motivations for the Support Vector Machine analysis in §4.2.

2.1.3. Known results on quiver mutation-acyclicity in rank 4

In rank 4, few existing results can help researchers determine if an arbitrary quiver Q is mutation-acyclic. There is currently no algorithm or mutation-invariant that can be used to decide mutation-acyclicity in ranks 4 and above, and there is reason to believe that no such algorithm exists for arbitrary quivers [24]. However, as we will see later, we find evidence of an underlying (possibly mutation-invariant) combinatorial structure that the machine can learn to distinguish mutation-acyclic and non-mutation-acyclic rank 4 quivers.

In what follows, we assume that Q only has one connected component; if Q is a rank 4 quiver that has more than one connected component, then the problem of mutation-acyclicity can be dealt with using the results in the previous section. The most applicable result in rank 4 (as well as higher ranks) is due to Buan, Marsh, and Reiten: Proposition 2.8.

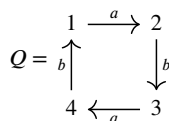
Definition 2.7. Let Q be a quiver on n vertices and $I \subseteq \{1, \dots, n\}$ be a subset of the vertex set. A quiver Q_I is a full subquiver of Q if its vertex set is $(Q_I)_0 = I$ and its arrows are given by $(Q_I)_1 = \{\alpha \in Q_1 \mid s(\alpha), t(\alpha) \in I\}$.

Proposition 2.8. [15, Corollary 5.3] Let Q be any quiver. If any full subquiver of Q is non-mutation-acyclic, then Q is non-mutation-acyclic.

In other words, if Q is a rank 4 quiver and contains a non-mutation-acyclic rank 3 quiver (a problem that has been dealt with in the previous

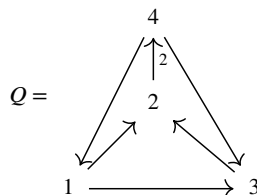
section), then Q itself is non-mutation-acyclic. This result is often useful in practice, especially when dealing with quivers of higher rank, since a rank n quiver contains $\binom{n}{3}$ rank 3 subquivers, and Markov constants are easy to calculate.

On the other hand, there are rank 4 quivers in which all (proper) full subquivers are mutation-acyclic, but the quiver itself is non-mutation-acyclic. In fact, in his dissertation, Warkentin [51] was able to provide an infinite family of examples of rank 4 quivers in which all of the proper subquivers are mutation-acyclic, but Q is non-mutation-acyclic. Specifically, he shows that quivers of the form



where $a, b \geq 2$ are non-mutation-acyclic. While the “box quivers” contain no rank 3 non-mutation-acyclic subquivers, they are sometimes mutation-equivalent to quivers Q' for which Proposition 2.8 can be applied.

There is also a well-known pathological quiver called the “dreaded torus” that is obtained from the triangulation of a torus with one boundary component and one marked point on the boundary:



The dreaded torus has the interesting property that mutation at any vertex produces an isomorphic copy of the dreaded torus quiver, making it effectively the only quiver in its mutation class (much like the Markov quiver). Since it contains an oriented cycle, it is non-mutation-acyclic. The existence of the dreaded torus (and possibly other pathological quivers) complicates the problem of deciding mutation-acyclicity using machines since Proposition 2.8 is not applicable.

The goal of this paper is to utilize computational techniques to assist researchers in classifying rank 4 quivers into mutation-acyclic and non-mutation-acyclic. To make theoretical progress on this problem, we show that the question is decidable for rank 4 quivers with at most 2 arrows between any pair of vertices.

Theorem 2.9. Suppose we have a quiver Q with 4 vertices, where the arrow weights take on values of 0, 1, or 2. Then Q is non-mutation-acyclic if and only if one of the following conditions holds.

- (1) Q contains a quiver in its mutation class that has a rank 3 non-mutation-acyclic subquiver.
 (2) Q is isomorphic to the dreaded torus quiver.

Proof. The full details of this computer-assisted proof are available on [GitHub](#); a summary of the logic is provided here. All quiver adjacency matrices for rank 4 quivers with arrow weights 0, 1, or 2 can be exhaustively generated, where the six entries of the upper triangle can take values from $\{-2, -1, 0, 1, 2\}$, hence producing $5^6 = 15625$ matrices. Of these, those that were not (weakly) connected, and hence not truly rank 4, were removed, leaving 15104 quivers. These adjacency matrices are still not unique up to graph isomorphism; therefore, a canonical form for the matrices was defined⁵ under this symmetry, and the set was reduced to 667 isomorphism classes of quiver adjacency matrices.

The representative quivers which are (strongly) acyclic are then trivially mutation-acyclic, 401 of the 667 were shown to be acyclic and were partitioned off into the mutation-acyclic side of the classification, leaving 266 unknown cases. Of these, 42 were shown to contain the Markov quiver as a subquiver, and two further cases were shown to be the 2-2-2-2 box quiver and dreaded torus. Partitioning these off resulted in 44 quivers on the non-mutation-acyclic side of the classification, with 222 remaining unknown cases. These 44 non-mutation-acyclic quivers were then each mutated exhaustively up to some set depth (8 in the case of this proof). Of the quivers produced in this way, 2 matched one of the 222 remaining unknown cases, which were then labeled as non-mutation-acyclic, leaving 220 unknown cases.

Each of these 220 unknown cases were then exhaustively mutated (using a breadth-first heuristic), producing their exchange graphs at ever greater depths until either an acyclic quiver was produced (and hence the initial quiver could be classified as mutation-acyclic) or a quiver which could be shown to be non-mutation-acyclic was produced (and hence the initial quiver could be classified as non-mutation-acyclic). Additionally, a flag was added that if the entire exchange graph was generated without producing an acyclic quiver, the initial quiver was marked as non-mutation-acyclic. The checks for non-mutation-acyclicity were carried out by searching each produced quiver's 3-vertex subquivers, and, when the subquiver was a directed cycle, performing the checks for the conditions (1)-(3) in Theorem 2.6; or by a produced quiver being isomorphic to a quiver from the generated exchange graphs seeded by each of the non-mutation-acyclic quivers in the previous step. These checks classified the remaining 220 quivers, using mutation depths up to 12.

Therefore, of the 667 isomorphism classes of rank 4 quivers with arrow weights 0, 1, or 2, the final class sizes (mutation-acyclic, non-mutation-acyclic) are (534, 133). These quivers are available on [GitHub](#), listed as their respective adjacency matrices in canonical form. \square

2.2. Machine learning

As a field, machine learning encompasses a broad range of computational techniques, often of statistical origin. Within ML, this set of techniques can be subdivided into three core subfields: supervised, unsupervised, and reinforcement learning.

Each subfield hosts methods suited to different data styles. Specifically, supervised learning deals with paired data such that each input has a corresponding output, and the techniques then address problems of function fitting and approximation, fitting a general, highly parameterized function to map from input to output. Unsupervised learning is in spirit data analysis, no output data is known for function fitting, so techniques are used to analyse the input dataset via methods of clustering and dimensional reduction. The final subfield of reinforcement learning deals with input data that may be represented as a state space,

where potential solutions to a problem can be evaluated with a score function, and then perturbed according to a set of actions.

The quiver data and problems considered here match most appropriately to methods in the first two subfields of supervised and unsupervised learning, and thus, in the following subsections, we introduce the techniques used from these areas in more detail.

2.2.1. Supervised learning

Depending on whether the output data in each data pair comes from a finite set or not, a supervised learning problem can be categorised into either classification or regression.

In this work, the aim is to identify the mutation-acyclic property. Therefore for input data as the adjacency matrix⁶ representation of a quiver, one can then assign a binary label to this quiver, which identifies whether it expresses the mutation-acyclic property or not. Therefore, the dataset takes the form $(B, 0)$ or $(B, 1)$ where $B = B(Q)$ is the skew-symmetric adjacency matrix of a quiver, and the binary output label indicates whether the quiver is mutation-acyclic. This output data is hence a finite set of size 2, making this problem a binary classification.

This dataset of pairs is partitioned into train data and test data, whereby the data is randomly shuffled, and 10% of the data is partitioned off to be used to independently test the model's performance after training has completed. Within the remaining train dataset, 30% of this is then also partitioned off to form the validation set, which allows independent tracking of the performance throughout training (whereas the test data is only used after training is complete). In practice, this process is generalised such that the dataset is partitioned 5 different ways (with the same split proportions but different shuffles of the full dataset prior to split), then 5 identically structured architectures are each trained and tested on one of these partitions. This process is cross-validation and produces five performance measures, which can be averaged to improve statistical confidence in the reported learning performance.

Since this is a classification problem, performance is assessed based on how frequently the trained model correctly identifies the true output class for the input datapoint. In this context, this is the ability of the model to identify whether the input quiver is mutation-acyclic or not. The test performance is recorded in a confusion matrix: C_{ij} , from which the standard learning performance metrics are derived. These metrics, along with their associated formulae, are provided in §A.1. Accuracy is used as the most standard interpretable metric, being the proportion of correctly classified test inputs, whilst Matthew's correlation coefficient (MCC) is used as a more stable, unbiased measure. Both measures evaluate to 1 for perfect learning, whereas for no learning, accuracy evaluates to 0.5 (for balanced binary classes) and MCC to 0.

During the training process, an optimiser algorithm will update the parameters of the model in order to minimise a loss function that evaluates the difference between the predicted and true output. The standard loss function for binary classification problems is binary cross-entropy, as used here with the formula provided in §A.1.

Neural networks Perhaps the most famous supervised learning architecture is that of neural networks (NNs). Largely due to their amenability to a variety of problems across many fields.

NNs are built from an array of neurons, where each neuron represents the application of a linear and then a non-linear function. For some input vector \underline{x} , the neuron acts as $\underline{x} \mapsto \phi(\underline{w} \cdot \underline{x} + b)$ for non-linear activation function ϕ and trainable parameters within the weight vector \underline{w} and bias number b . These neurons are then collected into layers, where every neuron in the layer receives the same input vector, being the concatenation of the all the previous layer's output numbers.⁷ To encourage

⁶ A quiver's 'adjacency matrix' is the same as its 'exchange graph', the former terminology is more common in data science literature, whereas the latter is the term used within the cluster algebras field.

⁷ The NN architecture described here is the prototypical dense feed-forward NN, which is quite general and reduces to other popular architectures such as

⁵ Using lexicographical ordering of the matrix elements.

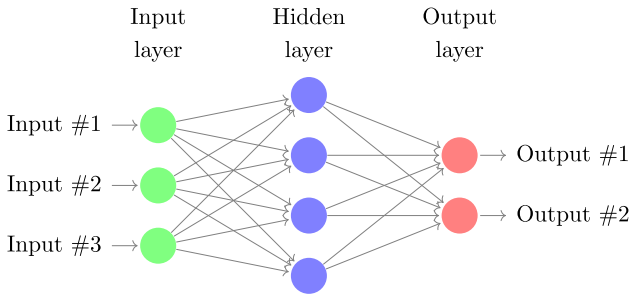


Fig. 2.2. An example neural network with one hidden layer; this can be generalised to more hidden layers and more neurons per layer. Each neuron represents the action of a linear and then a non-linear function on its input vector.

function stability during the training process and reduce the chances of overfitting, each neuron's input from a previous neuron can be chosen to be set to zero with a fixed probability, known as the dropout probability. The number of layers and number of neurons per layer are then hyperparameters which set the NN architecture, where the first layer neurons each receive the input vector (here the flattened adjacency matrix), and the output layer has as many neurons as the number of classes (for classification problems) and this layer's activation is set to a function which normalises the outputs to produce a probability distribution. An example diagram of a NN is given in Fig. 2.2.

The choice of activation function determines the structure of the function approximation; the most common choice is the rectified linear unit (ReLU), which sets the full NN function to be piecewise linear. There are generalisations of ReLU which smoothen the piecewise linear nature, including the scaled exponential linear unit (SELU), and the standard hyperbolic tangent function. Finally, for classification problems, where output normalisation to a probability distribution is desired, the softmax activation function is often used on the final layer (as chosen for the architectures in this work). This probability distribution over the output nodes is used in the loss for parameter updating, whilst for test classification, the output with the highest probability is selected as the predicted class for each test input datapoint. The formulae for these activation functions are provided and described in §A.1.1.

Over the training process, the optimiser updates the parameters (weights and biases) according to some adaptation of stochastic gradient descent, here Adam [37]. The train dataset inputs are shuffled and split into batches of a prespecified size, then they are fed into the NN to produce test outputs which the loss is evaluated at after differentiating it with respect to the model parameters. This process of iteratively updating parameters is repeated for all batches over the training dataset, completing one epoch. After this, the train dataset is shuffled and resplit into batches, and the whole process is repeated, now iterating over a fixed number of prespecified epochs. The step size in the parameter space taken by the optimiser with each update is scaled by a learning rate. The batch size, number of epochs, and learning rate are further hyperparameters one can choose to define a NN architecture.

Support vector machines Alternatively, perhaps the most optimised architectures for binary classification are support vector machines (SVMs). Their interpretability is a significant advantage for distilling true mathematical insight from their learning, and a feature we aim to take advantage of in this study.

Considering vectorial input data of size d , one can imagine each input as a point in \mathbb{R}^d , with the dataset then populating this space. For a binary classification problem, each datapoint is associated with a label

Convolutional NNs on zeroing of certain parameters. Furthermore, more exotic architectures can be created by allowing layers to input to themselves or to previous layers creating cycles in the NN graph (developing into recurrent NNs and transformers).

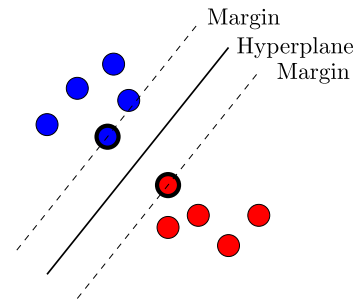


Fig. 2.3. An example support vector machine in 2-dimensions. The datapoints are coloured according to the class they belong to (either blue or red), and the support vectors used to construct the codimension-1 supporting hyperplane with maximum margin are bolded. Here, the hyperplane is linear, but the kernel trick could be used to construct higher-degree hyperplane equations.

indicating which of the two classes it belongs to. The SVM architecture then seeks to find a linear codimension-1 hyperplane within \mathbb{R}^d which best partitions the space such that either side of the plane corresponds to one of the classes. The performance is measured by the number of points on the correct side of the hyperplane for their label, and then the distance of the hyperplane to the closest points on either side (which, when maximised, improves the stability of generalisation).

Fig. 2.3 provides an example of a SVM fitting of 2-dimensional data, embedded in \mathbb{R}^2 , such that the codimension-1 linear hyperplane is a straight line. The data here is linearly separable, and hence, a hyperplane can be found that perfectly separates the data. In fact, one can see that rotating the hyperplane slightly would still separate the data perfectly, and thus, for optimum generalisability, one seeks the maximum-margin hyperplane. The margin is determined by orthogonally translating the hyperplane in both directions along the codimension-1 normal until the first point in each class is reached; the midpoint of this margin is the natural hyperplane position along its normal. These first points are the *support vectors* for the hyperplane, and completely determine it irrespective of the remaining datapoints. The training process then seeks the hyperplane orientation that maximises this margin, as demonstrated by the shown hyperplane position in the figure.

In reality, it is unlikely a generic dataset will be linearly separable; in that case, the hinge loss is used in place of it, defined as

$$\text{Hinge Loss} := \sqrt{\sum_i w_i^2} + C \left(\frac{1}{N} \sum_k \left(\max(0, 1 - y_k (\sum_i w_i x_{i,k}) - b) \right) \right) \quad (2.1)$$

for hyperplane equation $0 = \sum_i (w_i x_i) - b$ on input vectors $x_{i,k}$ with class labels y_k , where index i runs over the vector dimensions, and k the index of each (input vector, output label) in the dataset (of which there are N). The hyperplane equation parameters (w_i, b) are learned in the training process, where $2b/\sqrt{\sum_i w_i^2}$ determines the margin size. The use of the $\max(\cdot)$ function means data points that lie on the outside of the margins (with equations $\sum_i (w_i x_i) - b = \pm 1$) do not contribute to the loss as they are already correctly classified. The final parameter C dictates the regularisation, where this is large the main part of the loss function dominates (the soft-margin loss) and the optimiser prioritises a model which fits well with many datapoints correctly classified; conversely where C is low the regularisation first term in (2.1) dominates prioritising a simpler hyperplane equation with less parameters. Since the goal in the application of this architecture is interpretability, a low value of $C = 1$ is used.

Overall, the SVM parameters (w_i, b) are optimised throughout the training process in an iterative manner in order to minimise this hinge loss. The dataset is again split into train and test data, where the test data is reserved for use after training to evaluate performance.

This implementation of SVMs seeks an optimal linear hyperplane; however, many datasets are not linearly separable, where an optimal

non-linear hyperplane would be preferable. To facilitate the SVM process accommodating non-linear hyperplane equations, the ‘kernel trick’ is used. Here, the non-linear terms in the hyperplane equations are considered as additional dimensions, effectively trading the non-linear degrees of freedom for linear degrees of freedom in a higher-dimensional space. A linear hyperplane in this higher-dimensional space can then be mapped back to a non-linear hyperplane in the original space, and thus the ideology and process are very much the same.

The beauty in the trick is that the higher-dimensional embedding of each input is not required explicitly, saving substantial computational cost. Where the similarity between any two points in the higher-dimensional space is required for the training process, the kernel can compute this directly as a function of the original input vectors. For this application, we are primarily concerned with interpretability and stick to the simplest non-linear kernels, which are polynomial. For the fitting of a degree δ polynomial hyperplane, the kernel function form is:

$$\text{Ker}(\hat{x}_{k_1}, \hat{x}_{k_2}) := (\gamma(\hat{x}_{k_1} \cdot \hat{x}_{k_2}) + \beta)^\delta, \tag{2.2}$$

acting on input vectors \hat{x}_{k_1} or \hat{x}_{k_2} (where the dimension index is now implicit in the \cdot), using additional input parameters (γ, β) to optionally offset the hyperplane, here we stick to the default⁸ initialisations $\gamma = 1/(\sigma^2 d)$, $\beta = 0$; for data dimension d and data variance σ^2 .

Once a SVM has been trained, the hyperplane equation is given by

$$0 = \sum_k y_{\bar{k}} \alpha_{\bar{k}} \text{Ker}(\hat{x}_{\bar{k}}, \hat{x}) + b, \tag{2.3}$$

which can be evaluated for any input vector \hat{x} , and is based on the support vectors $\hat{x}_{\bar{k}}$ from the dataset (i.e. \bar{k} does not run over all dataset vectors, just those which are support vectors) with labels $y_{\bar{k}}$, explicitly using the kernel function for the specified degree as in (2.2), and the trained parameters⁹ $(\alpha_{\bar{k}}, b)$.

Using these supervised techniques, we aim to identify the mutation-acyclicity property in generic input quivers, particularly extracting full SVM equations to initiate interpretation.

2.2.2. Unsupervised learning

Depending on the goal of the data analysis, most unsupervised learning techniques can be applied for either dimensional reduction or clustering.

Prior to the application of supervised methods, it can be prudent to identify the optimal representation for the input data. In this work, we address this goal via the prototypical method of dimensional reduction on the quiver adjacency matrix data: *Principal Component Analysis* (PCA).

As a method, PCA utilises the diagonalisation of the dataset’s covariance matrix, where for the input dataset $\{\underline{x}\}$, the covariance K_{ij} is computed as

$$K_{ij} = E(x_i - E(x_i))(x_j - E(x_j)), \tag{2.4}$$

for x_i the i th component of the input vectors \underline{x} and $E(\cdot)$ the expectation value over the dataset. This matrix gives a centralised measure of second-order moments across the input components, and its diagonalisation produces eigenvectors, in this context called principal components, which determine a natural orthogonal basis for the data variation. Since the covariance matrix is symmetric, the eigenvalues are all real,¹⁰ and provide a means of determining which principal components are most informative on the data’s variation. Hence, for d -dimensional input vectors, the d principal components can be sorted in decreasing size of their eigenvalues.

⁸ Note for a linear kernel with $\delta = 1$ this kernel reduces to the simple dot product in the original input space (for $\gamma = 1, \beta = 0$).

⁹ These are alike the (w_j, b) in the linear case.

¹⁰ Actually, the covariance matrix is positive semi-definite, so the eigenvalues are also all non-negative.

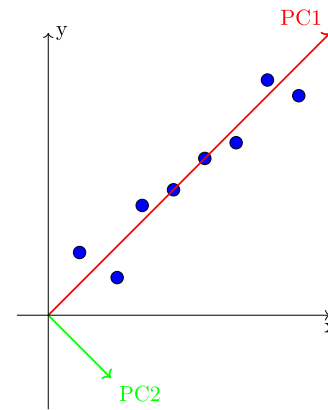


Fig. 2.4. An example of principal component analysis in 2-dimensions. The datapoints are coloured in blue, and after diagonalisation of the covariance matrix, the principal components (scaled by their respective eigenvalues) are also shown on the plot. From their respective sizes, projecting all datapoints onto PC1 would be an optimal 1-dimensional linear projection.

Therefore, to perform a dimensional reduction from the d -dimensional input onto some linearly optimum d' -dimensional subspace, one may project the dataset using a projection matrix formed from the first d' normalised principal components. The proportion of the data variation explained by this representation may be determined by the sum of the respective d' normalised eigenvalues.¹¹ A diagram demonstrating PCA decomposition of a 2-dimensional dataset is shown in Fig. 2.4, a linearly¹² optimum 1-dimensional projection would be along the first principal component.

3. Data generation

To study the properties of quivers, we must generate datasets that can be easily fed into statistical learning algorithms. In particular, recall that the operation of quiver mutation can be performed iteratively; thus, a starting point would be to generate datasets from exhaustively applying the mutation to a starting set of quivers. Methods of statistical learning are predominantly written in the language of linear algebra; hence, we must first express our quivers in this language. To this end, we represent every quiver in our dataset in terms of its exchange matrix b_{ij} , which for a rank n quiver has size $n \times n$. For computational convenience, we then flatten the components of our exchange matrix to a $1 \times n^2$ dimensional vector. In this paper, we restrict ourselves to rank 4 quivers, which correspondingly means that our dataset will be comprised of 1×16 dimensional vectors.

Example 3.1. The quiver in Fig. A.1a has adjacency matrix

$$\begin{pmatrix} 0 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \\ 0 & 0 & -1 & 0 \end{pmatrix}, \tag{3.1}$$

flattened to the input vector: $(0, 1, 0, 0, -1, 0, 1, 0, 0, -1, 0, 1, 0, 0, -1, 0)$. The respective output class label then depends on the investigation problem being learnt.

¹¹ Each normalised eigenvalue is often called the explained variance ratio. Here normalisation is across the spectrum of eigenvalues, whereas for each eigenvector it is within each vector.

¹² Note that PCA can be generalised to non-linear bases, also using kernels. However, here for this initial exploratory analysis, we focus on the simplest linear kernel.

Table 3.1

Dataset 1 Information.

Class	Number of Matrices	Class Label
A4-Like	39363	0
D4-Like	26242	1
NMA1	13121	2
NMA2	13121	3

Table 3.2

Dataset 2 Information.

Class	Number of Matrices	Class Label
MA	7285	0
NMA	2914	1

Table 3.3

Dataset 3 Information.

Class	Number of Matrices	Class Label
MA	236142	0
NMA	184785	1

Table 3.4

Dataset 4 Information.

Class	Number of Matrices
MA	12082
NMA	3022

For the three NN investigations carried out, we generate three datasets through this exhaustive mutation process: Dataset 1, Dataset 2, and Dataset 3.

Dataset 1 was generated from four cluster algebra mutation classes of quivers, named here: A4-like, D4-like, Non-Mutation-Acyclic-1 (NMA1), and Non-Mutation-Acyclic-2 (NMA2), whose initial quivers (defining the cluster algebra) are respectively illustrated in Figs. A.1, A.2, and A.3. For each initial class, we mutate the quivers to depth 8. The number of entries for each class (as well as the class labelling associated with each class) is given in Table 3.1.

Dataset 2 was similarly sourced from mutation classes with initial quivers depicted in Figs. A.1, A.2, and A.3. In this instance, we mutate each of the initial quivers up to depth 6. We then combine the A4-Like and D4-Like quivers into a single shuffled input class of mutation-acyclic (MA) quivers, and the NMA1 and NMA2 into a single shuffled input class of non-mutation-acyclic (NMA) quivers. The number of entries and class label is described in Table 3.2.

Dataset 3 is a larger equivalent of dataset 2, and is sourced from the exhaustive mutation of an initial set of mutation-acyclic quivers and non-mutation-acyclic quivers (see Table 3.3). For the mutation-acyclic class, we initially start off with a set of A4-like and D4-like quivers (as seen in A.1 and A.2, explicitly acyclic) with either 1, 2, 3, and 4 edges between the vertices. We then remove any isomorphic quivers from this initial set. To generate the full mutation set, we mutate each quiver in the initial quiver set up to mutation depth 7. Conversely, to generate the non-mutation-acyclic class, initial quivers given in Figs. A.4, A.5, and A.6 are used; with all the edges not a part of the 3-cycle in Fig. A.5 having lengths between 0 and 4. Then, removing any graphs that are not connected, and removing any isomorphic quivers. We then mutate the dataset up to depth 5.

The final investigation uses the more interpretable SVM architecture on a fourth dataset: Dataset 4. This dataset took the 667 rank 4 quivers with arrow weights $\{0, 1, 2\}$, classified according to being mutation-acyclic or non-mutation-acyclic via the proof of Theorem 2.9, and generated their full isomorphism classes (counts presented in Table 3.4).

Emphasising that the exponential nature of the mutation process makes data generation very unpredictable, where different mutation classes can have significantly different sizes at the same depth. The depth choice was hence motivated by a goal to generate sufficiently many quivers per class, taken to be ~ 10000 , and then consistently use that depth across the classes in an investigation. This is why different depths were used between investigation datasets. These datasets, as well as the generation scripts, are available at this paper's complementary [GitHub](#) repository.

4. Results

Using the datasets described in §3, machine learning methods are now implemented to learn the mutation-acyclic property for input quivers. Since quivers can be mutation-acyclic or not, this is, in general, a problem of binary classification; thus, popular classification architectures are selected for experimentation: the ubiquitous NN architecture and the more interpretable SVM architecture.

4.1. Neural networks: predicting mutation-acyclicity

As popular architectures for classification, NNs here are implemented for the task of identifying whether a generic input quiver is mutation-acyclic or not; a property extremely difficult to discern by eye (and generally analytically unknown for rank > 3 cases).

In doing so, three experiments are carried out, each a classification between classes of quivers sourced from a variety of cluster algebras, these datasets were described in §3, and generated from initial quivers depicted in §A.2. The experiments start by classifying between mutation classes, then between superclasses of many mutation classes chosen such that all quivers in the superclass are either mutation-acyclic or non-mutation-acyclic. The performance results are subsequently described.

4.1.1. Classifying mutation classes

As a warm-up exercise, this first experiment explores how effectively a NN can distinguish four different mutation classes of quivers, which we denote: A4-like, D4-like, Non-Mutation-Acyclic-1 (NMA1), and Non-Mutation-Acyclic-2 (NMA2); all contained within Dataset 1. This is similar in style to the NN classification in [20], albeit expanded to new mutation classes.

We utilize a fully connected deep neural network with 128 nodes in each layer with either a SELU, ReLU, or tanh activation function (NN structure given in Fig. A.8, established via heuristic hyperparameter tuning). Between each layer, we use a dropout rate of 0.4 to prevent overfitting. The final layer of the model had 4 nodes, which are activated by a softmax function, to output a normalised probability distribution over the classes. Dataset 1 is split into a train:test split of 90:10, with training data being split into a 70:30 split to accommodate the creation of a smaller validation training set. The model was trained over 10000 epochs with a learning rate of 0.0001 (using the Adam optimizer). In addition, each epoch utilized batching, using batch sizes of 200. The model was trained to minimize the Sparse Categorical Cross-entropy loss function. In addition, the dataset was balanced such that there was an equal distribution of each class in the dataset (for the smallest class size s , uniformly sample s quivers from each other class).

The model had an accuracy on test data of 61.8%, with a MCC of 0.507. These performances represent notable learning, since no ability to identify the cluster algebra from an input quiver is represented by (accuracy, MCC) scores of (25%, 0) respectively, which both of these substantially exceed. Since the quiver data between classes is impossible to classify by eye, and largely impractical to do analytically via mutation, these methods already display practical use in aiding the identification of cluster algebras for input quivers. These performance scores set a useful baseline for comparison in the subsequent experiments.

4.1.2. Classifying mutation-acyclicity: few algebras

In the second experiment, we tested how effective a NN was at determining the difference between mutation-acyclic quivers and non-mutation-acyclic quivers at rank 4. This experiment used Dataset 2 as its input, where two mutation classes of each type (mutation-acyclic or non-mutation-acyclic) were merged to form a larger database, causing the architecture to focus on the mutation-acyclicity property alone, instead of just the mutation class. Again, the data was class-balanced to ensure an equal distribution. The NN trained for this experiment is almost identical in structure to that used for the first experiment, with the only difference being that we have used 2 nodes in the output layer as opposed to 4 in experiment 1 of §4.1.1. Furthermore, the test:train split, validation split, learning rate, batch size, number of epochs, and optimizer are the same as those for experiment 1.

The model had an accuracy on test data of 87.7%, with a MCC of 0.764. Null learning for two outputs leads to (accuracy, MCC) performance scores of (50%, 0), and again, both the performances substantially exceed these values, demonstrating notable learning. The more informative MCC measure is particularly impressive, far exceeding 0, and indicating that the mutation-acyclicity property can be well identified for this database of quivers. This further hints at the existence of a combinatorial invariant of the quiver input data, which the NN function is approximating to then use for this classification.

4.1.3. Classifying mutation-acyclicity: many algebras

Following the promising results in §4.1.2, experiment 3 aimed to probe how the NN's ability to classify mutation-acyclic quivers and non-mutation-acyclic quivers would change under both a large increase in the data size and expanding the superclasses to include more cluster algebra mutation classes. To this end, we train the NN on the larger Dataset 3, where many more mutation classes of each type were used to form the classified mutation-acyclic/non-mutation-acyclic classes.

Owing to the increased size of the dataset, we adopt a shallower¹³ NN architecture to the previous two experiments (as shown in Fig. A.9). We utilize a fully connected deep neural network with 128 nodes in each layer with either a SELU or ReLU activation function. Between each layer, we use a dropout of 0.3 to avoid over-fitting. The final layer of the model had 2 nodes, which are activated by a softmax function. Dataset 3 is split into a train:test split of 90:10, with training data being further split into a 70:30 split to accommodate the creation of a smaller validation training set. The model was trained over 10000 epochs with a learning rate of 0.0001 (using the Adam optimizer). In addition, each epoch utilized batching, using batch sizes of 100. The model was again trained to minimize the Sparse Categorical Cross-entropy function. In addition, the dataset was balanced such that there was an equal distribution of each class in the dataset.

As the final NN experiment, to provide confidence in the learning performances, cross-validation was implemented, such that the experiment was run 5 times, with 5 independently initialised NNs, using random shuffling of the data before splitting and training within each run. The results for test accuracy and MCC across the runs, as well as the average, are outlined in Table 4.1. In addition, an averaged confusion matrix for predictions on the test set across all 5 runs is given in Fig. 4.1.

The results shown here express equally high performances to §4.1.2. They are also consistent across the cross-validation runs, providing statistical confidence in the average performance scores. Since the average MCC score far exceeds null performance ($0.747 \gg 0$, when $\text{MCC} \in [-1, 1]$), this strongly supports the existence of a mutation-invariant combinatorial constant that the NNs are approximating, which dictates

¹³ Since for this experiment we are primarily interested in the efficacy of the NN methods, the specific architecture is less important. Additionally, some experimentation with the previous architecture during hyperparameter tuning showed negligible performance differences despite longer training times.

Table 4.1

Results for Dataset 3 Training across the five cross-validation runs, displaying also the average performance scores with standard error.

Run	MCC	Test Accuracy Percentage
1	0.731	86.3%
2	0.754	87.4%
3	0.745	86.9%
4	0.748	87.1%
5	0.759	87.6%
Average	0.747 ± 0.004	$87.0 \pm 0.2\%$

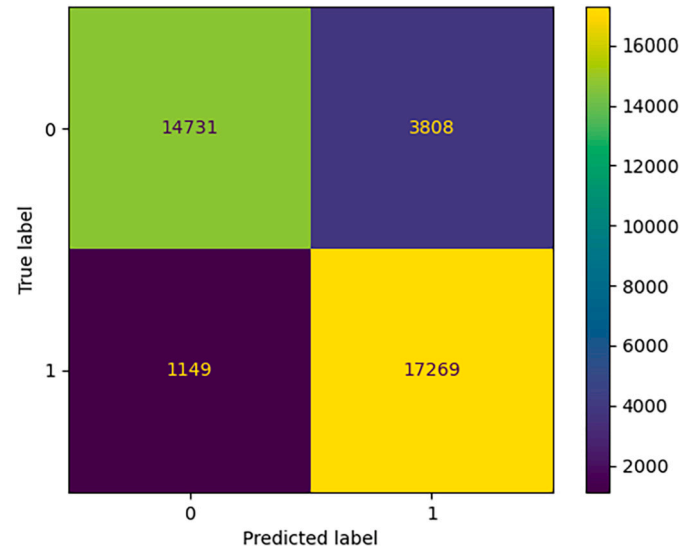


Fig. 4.1. Confusion matrix for predictions on Dataset 3's randomly sampled test set averaged across the five cross-validation runs for mutation acyclic quivers (0) and non-mutation acyclic quivers (1).

the mutation-acyclicity property, analogous to the Markov constant for rank 3 quivers. To open the door to its likely highly complex structure, we turn to a more interpretable ML method: SVMs, as outlined in the subsequent section.

4.2. Support vector machines: identifying separating hyperplanes

The results of §4.1 show promising performances, indicating there is likely some underlying structure in the edge multiplicity combinatorics that these highly parameterised NN models can leverage. However, as highly non-linear models, they are difficult to interpret and to extract mathematical insight from.

In this section, we shift focus to a complementary data setup (i.e., splitting into mutation-acyclic and non-mutation-acyclic), but now considering all possible quivers with an edge multiplicity < 3 . This is Dataset 4, described in §3, from the proof of Theorem 2.9. Additionally, we use a more interpretable ML architecture: SVMs (as introduced in §2.2.1), with the intention that the simpler design and directly extractable equations will be more insightful to mathematicians seeking to design appropriate general invariants of mutation-acyclicity.

4.2.1. Principal component analysis

To develop an analytic understanding of the separation between NMA and MA quivers, it is advantageous to first determine how each type of quiver is distributed within the dataset (Dataset 4). Recalling that each quiver is represented by a 1×16 vector, we see that our dataset is defined within a 16-dimensional space, which may be taken to be \mathbb{R}^{16} . For computational efficiency, it would be effective to probe which of the coordinate directions the dataset is either aligned with or most spread across. In addition, it would also be prudent to identify any patterns

Table 4.2
Eigenvalues for PCA transformed rank 4 quiver dataset (Dataset 4), those $< 10^{-30}$ have been rounded to 0.

	Eigenvalue
1	0.166666666666666674
2	0.166666666666666674
3	0.166666666666666669
4	0.166666666666666669
5	0.166666666666666663
6	0.166666666666666657
7,8,...,16	0

or correlations between the various coordinate directions. The overall aim of this procedure would be to find which directions are most significant and then project our dataset to only depend on these favourable coordinate directions. PCA, as introduced in §2.2.2, provides a useful tool in this instance to determine which of 16 dimensions are the most pivotal in describing the overall nature of the data, and what efficient compressed representations exist.

In performing PCA (with a linear kernel) on the dataset, we allow for the choice of a generic basis for which the dataset is transformed (which is an artifact of the algorithm e.g. depending on the order of the dataset) for this analysis. The eigenvalues of the PCA decomposition for this new generic basis are given in Table 4.2. From Table 4.2, it becomes clear that the first 6 dimensions (which share the same eigenvalue) are the most important for our analysis, as they are by far the largest eigenvalues we have for all 16 dimensions.

In Fig. 4.2, we plot the spread of the dataset pairwise between each of these 6 dimensions. The above results imply that we can form an eigenbasis from the 6 most important directions, onto which we can project our dataset, to reduce its dimensionality (from 16 to 6).

However, before we proceed any further, we observe that the exchange matrix for a rank 4 quiver is:

$$b_{ij} = \begin{pmatrix} 0 & e_0 & e_1 & e_2 \\ -e_0 & 0 & e_3 & e_4 \\ -e_1 & -e_3 & 0 & e_5 \\ -e_2 & -e_4 & -e_5 & 0 \end{pmatrix}, \tag{4.1}$$

due to the representation’s antisymmetry, where e_0 through to e_5 are integers. Therefore, these input exchange matrices have 6 degrees of freedom, and can naturally be represented in a 6-dimensional space as the vector: $(e_0, e_1, e_2, e_3, e_4, e_5)$.

The PCA eigenvalues nicely corroborate this 6-dimensional representation, confirming that no further sensible linear compression is possible, and since the method is linear, the PCA eigenvectors are some linear combination of these values. Therefore, motivated by the PCA, the next investigations will use this lossless compressed 6-dimensional representation as the quiver input.

4.2.2. SVM analysis

With a more natural and efficient representation of the input quiver data, we can now train a SVM, as introduced in §2.2.1, on the dataset to identify a 5-dimensional (codimension-1) separating hyperplane between the mutation-acyclic (MA) and non-mutation-acyclic (NMA) quivers in our dataset. To maximise interpretability, the SVM analysis was initially applied with a linear kernel, then generalised with polynomial kernels of increasing degree, covering orders 1 to 26. In order to classify the dataset (of Dataset 4) into either MA or NMA, we assigned MA quivers a label of -1 , and NMA quivers the label of 1 . Moreover, we assigned class weights, with MA quivers having a weight of 1 and NMA quivers having a weight of 3.998 . This was done to reflect that NMA quivers had a smaller population than MA in the dataset, and this weighting for SVM training exactly inverts this proportion such that both classes are viewed with equal importance in the training loss. Additionally, the reg-

Table 4.3
MCC of the trained SVM hyperplanes with varying degree polynomial kernels, reported to 3 decimal places; shown with the hyperplane equation number of terms.

Polynomial Kernel Degree	MCC	Number of Terms
1	0.000	8
2	0.410	22
3	0.000	58
4	0.807	128
5	0.000	254
6	0.879	464
8	0.907	1288
10	0.916	3004
12	0.921	6189
14	0.921	11629
16	0.921	20350
18	0.924	33650
20	0.909	53132
22	0.924	80740
24	0.868	118800
26	0.868	170044

ularisation parameter C was set to 1 , leading to a higher regularisation and prioritising identifying a simpler, more interpretable hyperplane.

We outline the learning results for each kernel order on the data in Fig. 4.3 and Table 4.3. Here, we take MCC to be the test accuracy measure, as it is better adapted to the uneven class sizes.

From Table 4.3, we see that odd ordered polynomials kernels prove completely ineffective in separating NMA from MA; likely since the data naturally contains all quivers in each relabelling isomorphism orbit, so is naturally symmetric, which the odd degree functions cannot accommodate. However, even ordered polynomial kernels prove more successful in separating the classes. In particular, the order 18 polynomial from the degree 18 kernel exhibited the highest classification performance between NMA and MA quivers, with a MCC of 0.924 .

The MCC performance results for each polynomial kernel are listed in Table 4.3, with the number of terms in each polynomial equation. Additionally, the analytic equations for the identified optimally-classifying SVM hyperplanes are available at the paper’s [GitHub](#).

At this juncture, we also note that additional SVM training was performed on the so-called canonical form of Dataset 4 (isomorphism classes). This resulted in, at face value, favourable training results (an MCC score of 0.97 at SVM polynomial kernel order 12); however, the number of terms in the resultant classification polynomials exceeded the number of degrees of freedom for the system; hence, they are not reported here as memorisation schemes are likely occurring.

The equation determined from the order 6 polynomial is the most interesting to us, as it exhibits excellent prediction despite a low kernel degree and number of terms.¹⁴ Since Dataset 4 contains 15104 quivers, and respectively 90624 input integers, despite these equations being long (where the degree 6 equation has 464 terms), this is still orders of magnitude less degrees of freedom than the input dataset; such that the SVMs are consolidating the mutation-acyclic property quite significantly. The existence of an equation that can classify so well solidifies this motivation behind the existence of a mutation invariant for rank 4 quivers, which determines the mutation-acyclicity property. Despite this degree 6 equation perhaps still being over-parameterised, the explicit nature of the equation’s representation causes it to be a combinatorial invariant, with the hope that the equation provides some hints towards a more intuitive construction.

¹⁴ The regularisation contribution could have been increased to discourage superfluous terms in the SVM equation, however this shifts the learning goal away from the core problem and the interpretation of the near-zero coefficients of non-linear basis monomials is somewhat ambiguous. Hence, focus on a smaller, more interpretable basis is desirable.

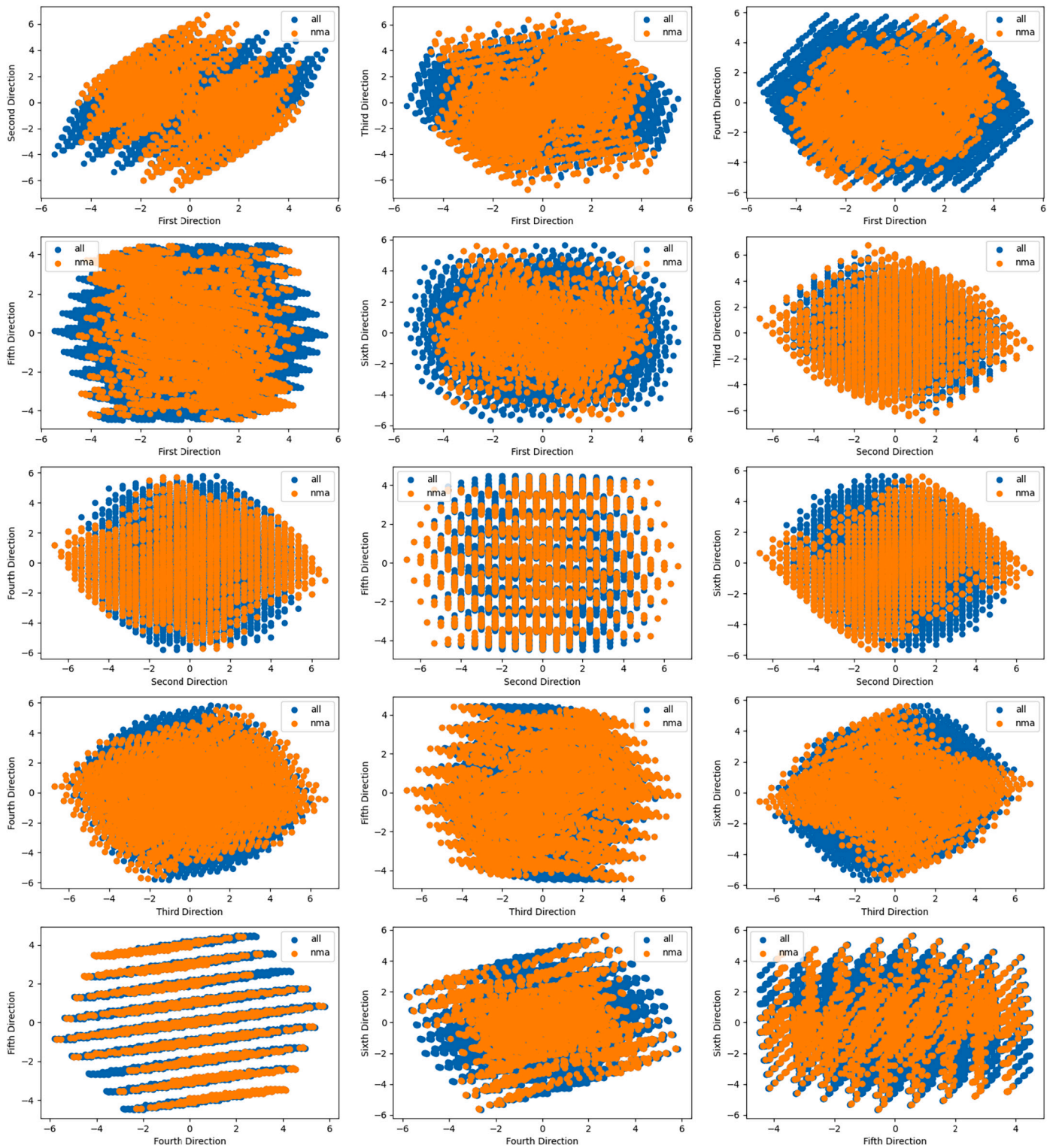


Fig. 4.2. Spread of non-mutation-acyclic quivers (orange dots) vs mutation-acyclic quivers (blue dots) for pairwise projections of the 6 principal directions chosen from PCA of Dataset 4.

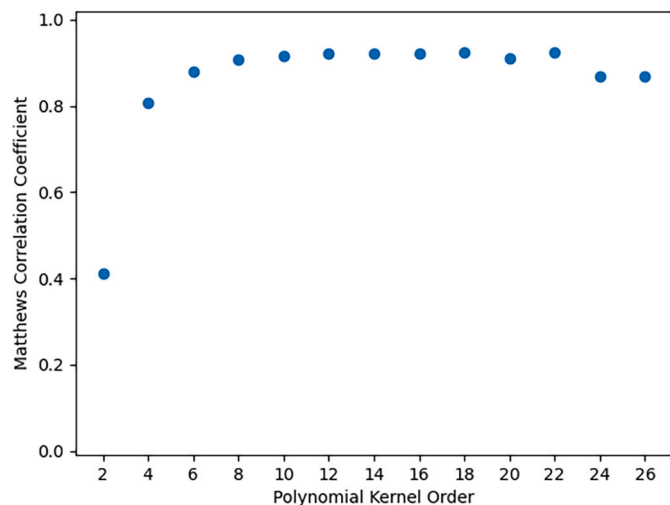


Fig. 4.3. Final MCC scores for Support Vector Machines with even polynomial kernels of increasing degree, trained on the compressed 6-dimensional input for rank 4 quivers with edge multiplicity < 3 .

Higher edge multiplicities As a final remark on these results, we note again that the analytic expression for the separating hyperplanes was derived from a dataset that contained rank 4 quivers with either 0, 1, or 2 edges between each vertex. An immediate question to ask is how well the hyperplane generalises to classify rank 4 quivers with edge multiplicities ≥ 3 .

In testing this, we trial the machine learnt degree 18 polynomial kernel equation on a selection of quivers with higher edge multiplicities; as shown in §A.2.1, all of which are classified correctly. These quivers include the A4-like type (Fig. A.7a) and NMA type (Fig. A.7d) edited to include edges with multiplicity ≥ 3 , as well as a quiver for each of these reached after a random sequence of mutations (Figs. A.7b, A.7e), finally each of these types is also shown with much larger edge multiplicities (Figs. A.7c, A.7f).

On this paper’s respective [GitHub](#) repository a script is also available to test any quiver of the reader’s choosing for mutation-acyclicity using this trained degree-18 SVM; simply enter the adjacency matrix of the quiver and the script will output a prediction. How the performance extends in general beyond Dataset 4, to higher edge multiplicities, and how the mutation invariant defined by the trained SVM polynomial can be generalised in these cases is left to future work.

5. Conclusion

In this paper, we explored the potential for ML techniques to tackle a complex problem within quiver mutation, specifically in detecting mutation-acyclicity. Our experiments yielded strong results, with NNs and SVMs achieving high accuracy in distinguishing mutation-acyclic quivers from their non-mutation-acyclic counterparts.

While these findings affirm the efficacy of ML in this domain, they also open new avenues for further investigation. The explicit equation formulations learnt by the SVMs of each kernel degree (available on [GitHub](#)) aim to provide a starting basis for distilling mathematical insight about the design of mutation-invariant combinatorial constants that identify mutation-acyclicity, alike the Markov constant, for higher rank quivers. This interpretative phase will be crucial for translating the empirical success of ML into theoretical advancements.

Additionally, the successes of these ML investigations motivated the computational work in proving the new Theorem 2.9, another central result of this paper. It is worth emphasising in the previously classified rank 3 case of Theorem 2.6 (2), the exceptional cases for identifying mutation-acyclicity all have edge weights ≤ 2 , as has been considered for this Theorem in the rank 4 case. The new Theorem’s low edge-weight

results for rank 4, hence hope to be useful for future work for a more general result by providing an exhaustive check of these likely exceptional cases, as well as to be useful in practical work where academics often use low edge-weight quivers in cluster algebra seeding.

In summary, this work not only establishes a solid foundation for the use of ML in quiver mutation analysis but also provides additional evidence for the exciting possibilities that lie ahead in integrating data-driven approaches with traditional mathematical research. By continuing to embrace this ML framework, we aim to deepen our understanding of mathematical structures and contribute to the broader intersection of machine learning and mathematics research.

CRediT authorship contribution statement

Kymani T.K. Armstrong-Williams: Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Edward Hirst:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Blake Jackson:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization. **Kyu-Hwan Lee:** Writing – review & editing, Writing – original draft, Visualization, Validation, Supervision, Software, Resources, Project administration, Methodology, Investigation, Funding acquisition, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Kymani Armstrong-Williams reports financial support was provided by Science and Technology Facilities Council. Kyu-Hwan Lee reports financial support was provided by Simons Foundation. Kymani Armstrong-Williams reports financial support was provided by Engineering and Physical Sciences Research Council. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

KAW is supported by a studentship from the UK Engineering and Physical Sciences Research Council (EPSRC) EP/T518086/1 and by the UK Science and Technology Facilities Council (STFC) Consolidated Grant ST/P000754/1 “String theory, gauge theory and duality”. EH acknowledges support from Pierre Andurand over the course of this research. KHL’s work was partially supported by a grant from the Simons Foundation (#712100). This research utilised Queen Mary’s Apocrita HPC facility [36], supported by QMUL Research-IT.

Appendix A

A.1. Machine learning metrics

Test performances in supervised classification problems are recorded in a confusion matrix: C_{ij} . In the binary classification scenario, this is a 2×2 matrix where the row index is the true class of the input and the column index is the model-predicted class of the input. As the test data is run through the model, the respective C_{ij} entry is incremented according to each test input’s true and predicted class; then, after testing is complete, the matrix is appropriately normalised. All popular classification performance measures are then computed as functions of this

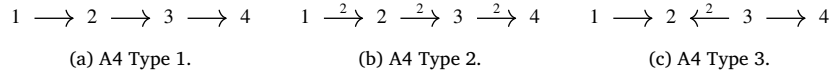


Fig. A.1. A4-like Quivers.

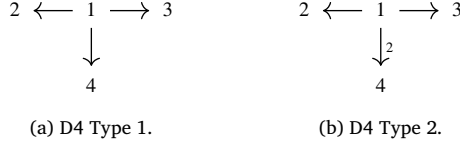


Fig. A.2. D4-like Quivers.

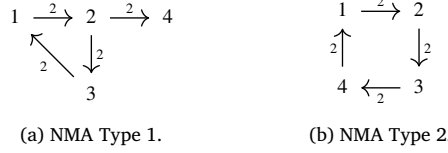


Fig. A.3. NMA Quivers.

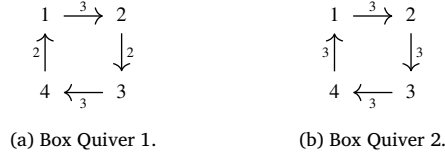


Fig. A.4. NMA Quivers Box Type.

matrix, the most common being accuracy, which is just the proportion of correctly classified inputs and is equal to the matrix trace. However, this measure is prone to bias where the true classes are unbalanced; therefore, a more appropriate measure, which includes contribution from the off-diagonal components (the statistical Type I and Type II errors), is Matthew’s correlation coefficient (MCC), defined, along with accuracy, as:

$$\text{Accuracy} := \sum_i C_{ii} = C_{00} + C_{11}, \tag{A.1}$$

$$\text{MCC} := \frac{C_{11}C_{00} - C_{01}C_{10}}{\sqrt{(C_{11} + C_{01})(C_{11} + C_{10})(C_{00} + C_{01})(C_{00} + C_{10})}}, \tag{A.2}$$

both measures evaluate to 1 for perfect learning, whereas for no learning, accuracy evaluates to 0.5 (for balanced binary classes) and MCC to 0.

In these classification problems, the output is often a discrete probability distribution over the possible classes, from which the loss function used in training, binary cross-entropy, is defined as:

Binary Cross-Entropy

$$:= \frac{1}{N} \sum_i (p_{true=0,i} \log(p_{pred=0,i}) + p_{true=1,i} \log(p_{pred=1,i})), \tag{A.3}$$

where the average is taken over the training batch of size N , using the true probabilities $p_{true=\#,i}$ of the output being in either class 0 or 1 (which as the data is labelled is exactly 1 for one probability and 0 for the other), and the predicted probabilities $p_{pred=\#,i}$ (satisfying $1 = p_{pred=0,i} + p_{pred=1,i} \forall i$).

A.1.1. Neural network activations

Fundamental to the structure of neural network architectures is their non-linearity, introduced by the choice of an activation function. The standard options, as used in the architectures in this work, are detailed below.

The choice of activation function determines the structure of the function approximation; the most common choice is the rectified linear unit (ReLU), defined:

$$\phi_{\text{ReLU}}(x) := \begin{cases} x & x \geq 0, \\ 0 & \text{otherwise.} \end{cases} \tag{A.4}$$

This choice sets the full NN function to be piecewise linear, and then it is this function style used to approximate the true map connecting the input to the output spaces. There are generalisations of ReLU which smoothen the piecewise linear nature, a candidate used here is the scaled exponential linear unit (SELU), defined:

$$\phi_{\text{SELU}}(x) := \begin{cases} \lambda x & x \geq 0, \\ \lambda \alpha (e^x - 1) & \text{otherwise,} \end{cases} \tag{A.5}$$

for constants (λ, α) taking their standard values (1.05070098, 1.67326324). Another common activation also used is the standard hyperbolic tangent function, $\phi_{\text{tanh}}(x) = \tanh(x)$. For the final output layer, where output normalisation to a probability distribution is desired, the softmax activation function is used:

$$\phi_{\text{softmax}}(x_i) := \frac{e^{x_i}}{\sum_j e^{x_j}}, \tag{A.6}$$

acting on the i th node output in the output layer x_i , normalising using a sum over all j nodes in the output layer.

A.2. Initial quivers

Here, the initial quivers used to seed the generation of quivers in the mutation classes considered are shown. The generated mutation classes of quivers were then combined in various ways to form the datasets, described in §3, used in the ML investigations of §4.

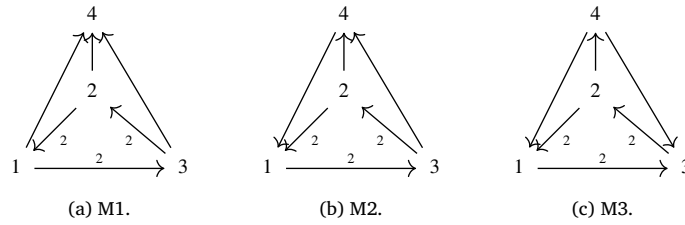


Fig. A.5. NMA-like Graphs Set 1.

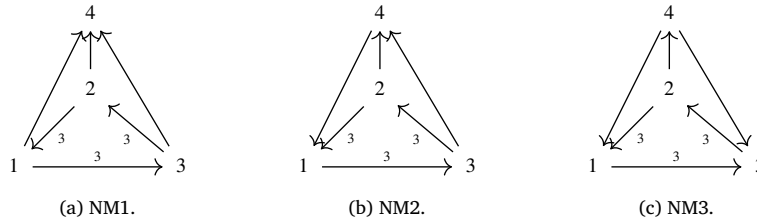


Fig. A.6. NMA-like Graphs Set 2.

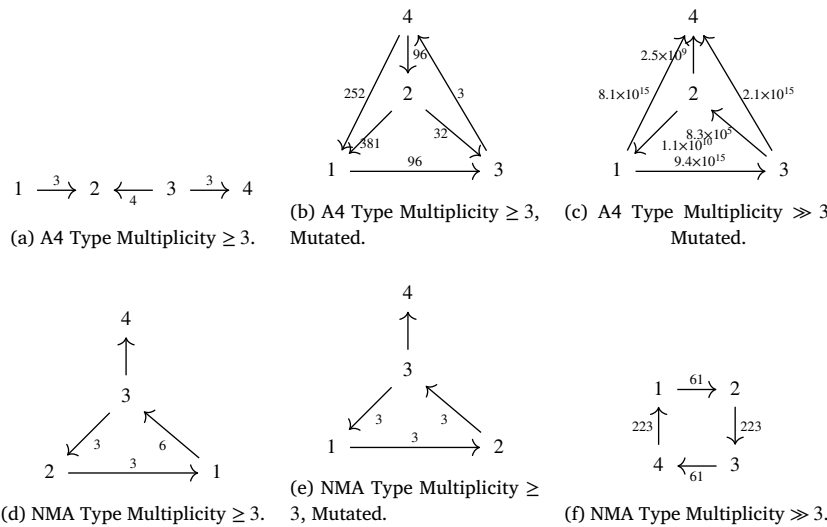


Fig. A.7. Quivers classified correctly by the trained degree-18 SVM polynomial.

A.2.1. High edge multiplicity quivers

A selection of quivers with higher-multiplicity edges is shown below, all of which are correctly classified by the trained degree-18 SVM polynomial.

A.3. NN architecture diagrammatics

Here are flowcharts displaying the NN architecture setups used in the learning of §4.1. Fig. A.8 is the architecture used for the experiments for §4.1.1 and §4.1.2; whereas Fig. A.9 is for §4.1.3.

Data availability

The datasets generated during the current study are shared at GitHub (https://github.com/KTKAW/MACHINE_LEARNING_MUTATION_ACYCLICITY_OF QUIVERS.git).

References

[1] Martín Abadi, Ashish Agarwal, Paul Barham, Eugene Brevdo, Zhifeng Chen, Craig Citro, Greg S. Corrado, Andy Davis, Jeffrey Dean, Matthieu Devin, Sanjay Ghemawat,

Ian Goodfellow, Andrew Harp, Geoffrey Irving, Michael Isard, Yangqing Jia, Rafal Jozefowicz, Lukasz Kaiser, Manjunath Kudlur, Josh Levenberg, Dandelion Mané, Rajat Monga, Sherry Moore, Derek Murray, Chris Olah, Mike Schuster, Jonathon Shlens, Benoit Steiner, Ilya Sutskever, Kunal Talwar, Paul Tucker, Vincent Vanhoucke, Vijay Vasudevan, Fernanda Viégas, Oriol Vinyals, Pete Warden, Martin Wattenberg, Martin Wicke, Yuan Yu, Xiaoqiang Zheng Tensor, Flow: large-scale machine learning on heterogeneous systems, Software available from [tensorflow.org](https://www.tensorflow.org), 2015.

[2] Daattavya Aggarwal, Yang-Hui He, Elli Heyes, Edward Hirst, Henrique N. Sá Earp, Tomás S.R. Silva, Machine learning Sasakian and G2 topology on contact Calabi-Yau 7-manifolds, *Phys. Lett. B* 850 (2024) 138517.

[3] Rashid Alawadhi, Daniele Angella, Andrea Leonardo, Tancredi Schettini Gherardini, Constructing and machine learning Calabi-Yau five-folds, *Fortschr. Phys.* 72 (2) (2024) 2300262.

[4] Malik Amir, Yang-Hui He, Kyu-Hwan Lee, Thomas Oliver, Eldar Sultanow, Machine Learning Class Numbers of Real Quadratic Fields, 2022.

[5] Guillermo Arias-Tamargo, Yang-Hui He, Elli Heyes, Edward Hirst, Diego Rodriguez-Gomez, Brain webs for brane webs, *Phys. Lett. B* 833 (2022) 137376.

[6] Anthony Ashmore, Yang-Hui He, Elli Heyes, Burt A. Ovrut, Numerical spectra of the Laplacian for line bundles on Calabi-Yau hypersurfaces, *J. High Energy Phys.* 07 (2023) 164.

[7] Ibrahim Assem, Martin Blais, Thomas Brüstle, Audrey Samson, Mutation classes of skew-symmetric 3×3 -matrices, *Commun. Algebra* 36 (4) (2008) 1209–1220.

[8] Ibrahim Assem, Daniel Simson, Andrzej Skowroński, Elements of the representation theory of associative algebras. Vol. 1, in: *Techniques of Representation Theory*, in:

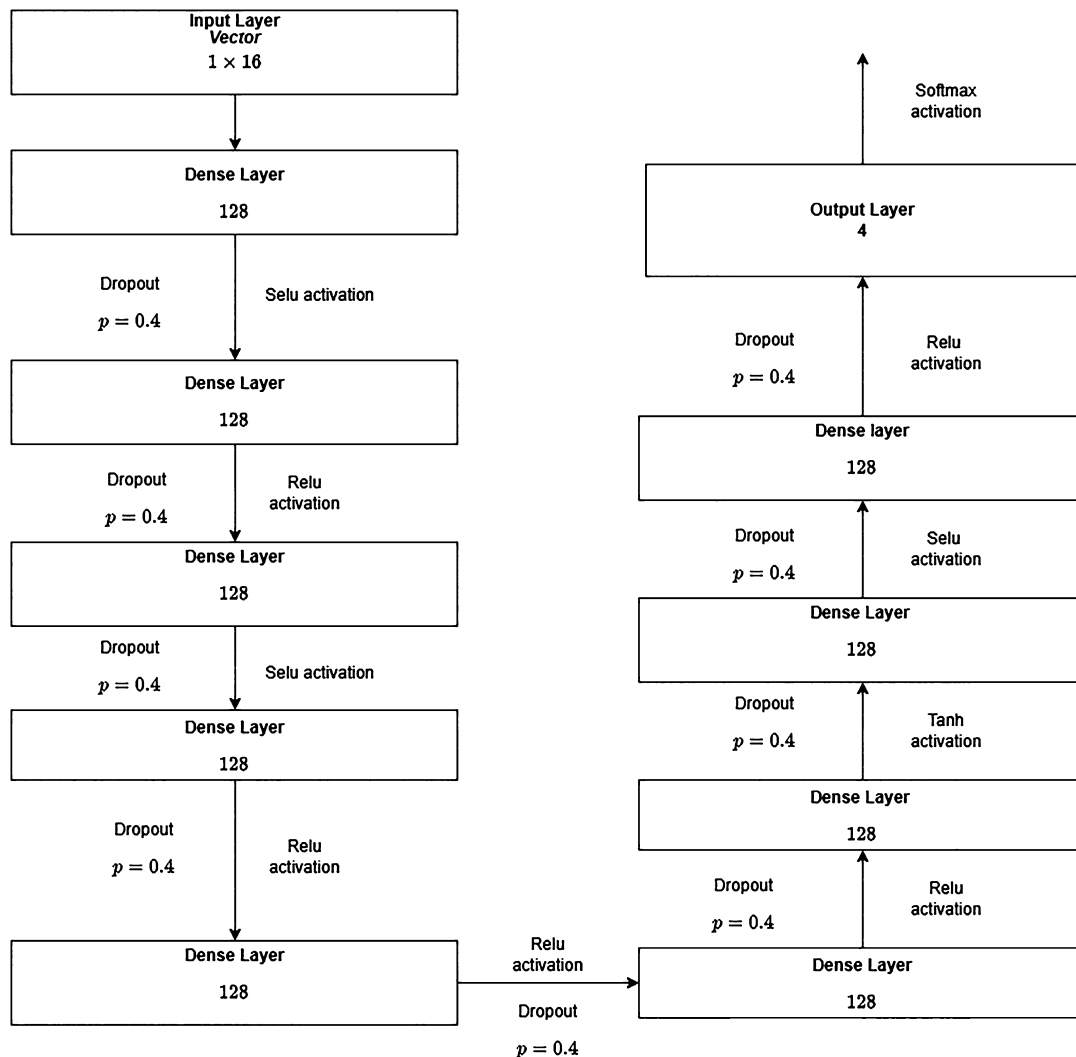


Fig. A.8. Diagram of the NN architecture used in experiments §4.1.1 and §4.1.2.

London Mathematical Society Student Texts, vol. 65, Cambridge University Press, Cambridge, 2006.

[9] Jiakang Bao, Sebastián Franco, Yang-Hui He, Edward Hirst, Gregg Musiker, Yan Xiao, Quiver mutations, Seiberg duality and machine learning, *Phys. Rev. D* 102 (8) (2020) 086013.

[10] Jiakang Bao, Yang-Hui He, Edward Hirst, Neurons in amoebae, *J. Symb. Comput.* 116 (2023) 1–38.

[11] Andre Beineke, Thomas Brüstle, Lutz Hille, Cluster-cyclic quivers with three vertices and the Markov equation, *Algebr. Represent. Theory* 14 (1) (2011) 97–112. With an appendix by Otto Kerner.

[12] Per Berglund, Yang-Hui He, Elli Heyes, Edward Hirst, Vishnu Jejjala, Andre Lukas, New Calabi–Yau manifolds from genetic algorithms, *Phys. Lett. B* 850 (2024) 138504.

[13] David S. Berman, Yang-Hui He, Edward Hirst, Machine learning Calabi-Yau hypersurfaces, *Phys. Rev. D* 105 (6) (2022) 066002.

[14] Klaus Bongartz, Algebras and quadratic forms, *J. Lond. Math. Soc. (2)* 28 (3) (1983) 461–469.

[15] Aslak Bakke Buan, Bethany R. Marsh, Idun Reiten, Cluster mutation via quiver representations, *Comment. Math. Helv.* 83 (1) (2008) 143–177.

[16] Pietro Capuozzo, Tancredi Schettini Gherardini, Benjamin Suzzoni, Machine Learning Toric Duality in Brane Tilings, 2024.

[17] Siqu Chen, Yang-Hui He, Edward Hirst, Andrew Nestor, Ali Zahabi, Mahler measuring the genetic code of amoebae, *Adv. Theor. Math. Phys.* 27 (5) (2023) 1405–1461.

[18] Man-Wai Cheung, Pierre-Philippe Dechant, Yang-Hui He, Elli Heyes, Edward Hirst, Jian-Rong Li, Clustering Cluster Algebras with Clusters, 2022.

[19] Lyndie Chiou, Elliptic curve ‘murmurations’ found with AI take flight, *Qunata Magazine* (2024).

[20] Pierre-Philippe Dechant, Yang-Hui He, Elli Heyes, Edward Hirst, Cluster algebras: network science and machine learning, *J. Comput. Algebra* 8 (2023).

[21] Tucker J. Ervin, Blake Jackson, Kyungyong Lee, Son Dang Nguyen, Geometry of C -vectors and C -Matrices for Mutation-Infinite Quivers, arXiv:2410.08510 [math.CO], October 2024.

[22] Anna Felikson, Pavel Tumarkin, Acyclic cluster algebras, reflection groups, and curves on a punctured disc, *Adv. Math.* 340 (2018) 855–882.

[23] Bo Feng, Amihay Hanany, Yang-Hui He, Angel M. Uranga, Toric duality as Seiberg duality and brane diamonds, *J. High Energy Phys.* 2001 (12) (Dec 2001) 035.

[24] Sergey Fomin, Scott Neville, Long mutation cycles, arXiv:2304.11505 [math], April 2023.

[25] Sergey Fomin, Andrei Zelevinsky, Cluster algebras. I. Foundations, *J. Am. Math. Soc.* 15 (2) (2002) 497–529.

[26] Mark Gross, Paul Hacking, Sean Keel, Maxim Kontsevich, Canonical bases for cluster algebras, *J. Am. Math. Soc.* 31 (2) (2018) 497–608.

[27] Aric A. Hagberg, Daniel A. Schult, Pieter Swart, Exploring network structure, dynamics, and function using networkx, 2008.

[28] Baran Hashemi, Roderic G. Corominas, Alessandro Giacchetto, Can Transformers do Enumerative Geometry?, 2024.

[29] Yang-Hui He, Kyu-Hwan Lee, Thomas Oliver, Machine-learning number fields, *Comput. Geom. Data* 2 (2022) 49–66.

[30] Yang-Hui He, Kyu-Hwan Lee, Thomas Oliver, Machine-learning the Sato–Tate conjecture, *J. Symb. Comput.* 111 (2022) 61–72.

[31] Yang-Hui He, Kyu-Hwan Lee, Thomas Oliver, Machine-learning invariants of arithmetic curve, *J. Symb. Comput.* 115 (2023) 478–491.

[32] Yang-Hui He, Kyu-Hwan Lee, Thomas Oliver, Alexey Pozdnyakov, Murmurations of elliptic curves, *Exp. Math.* (2024).

[33] Edward Hirst, Tancredi Schettini Gherardini, Calabi-Yau four-, five-, sixfolds as Pwn hypersurfaces: machine learning, approximation, and generation, *Phys. Rev. D* 109 (10) (2024) 106006.

[34] V.G. Kac, Infinite root systems, representations of graphs and invariant theory, *Invent. Math.* 56 (1) (1980) 57–92.

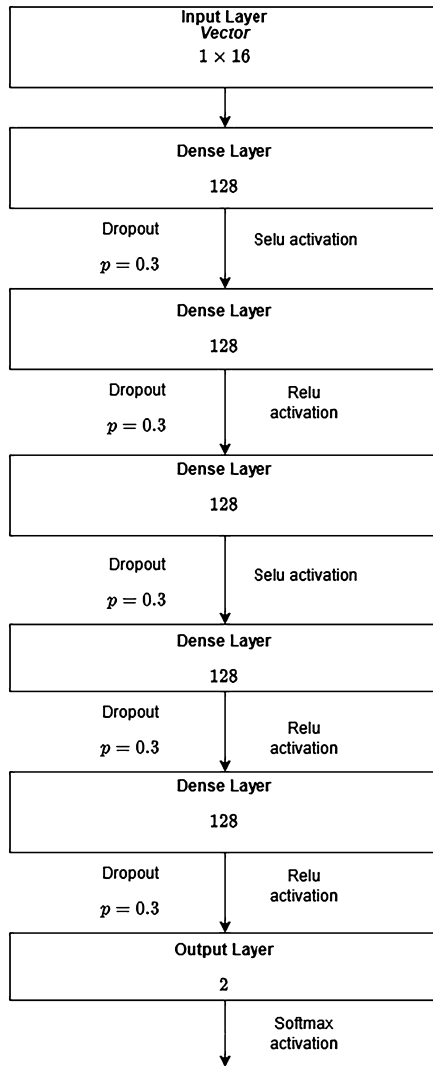


Fig. A.9. Diagram of the NN architecture used in experiment §4.1.3.

[35] Bernhard Keller, Quiver mutation and combinatorial DT-invariants, September 2017.

[36] Thomas King, Simon Butcher, Lukasz Zalewski, Apocrita - High Performance Computing, Cluster for Queen Mary University of London, March 2017.

[37] Diederik P. Kingma, Jimmy Ba Adam, A method for stochastic optimization, 2017.

[38] Maxim Kontsevich, Yan Soibelman, Stability structures, motivic Donaldson-Thomas invariants and cluster transformations, November 2008.

[39] Kyu-Hwan Lee, Kyungyong Lee, A correspondence between rigid modules over path algebras and simple curves on Riemann surfaces, Exp. Math. 30 (3) (2021) 315–331.

[40] Kyu-Hwan Lee, Kyungyong Lee, Matthew R. Mills, Geometric description of C-vectors and real Lösungen, Math. Z. 303 (2) (2023) 44.

[41] Greg Muller, Locally acyclic cluster algebras, Adv. Math. 233 (2013) 207–247.

[42] Greg Muller, $A = U$ for locally acyclic cluster algebras, SIGMA 10 (2014) 094.

[43] Gregg Musiker, Christian Stump, A compendium on the cluster algebra and quiver package in sage, 2011.

[44] Alfredo Nájera Chávez, On the c-vectors of an acyclic cluster algebra, Int. Math. Res. Not. 6 (2015) 1590–1600.

[45] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, E. Duchesnay, Scikit-learn: machine learning in python, J. Mach. Learn. Res. 12 (2011) 2825–2830.

[46] Tiago P. Peixoto, The graph-tool python library. figshare, 2014.

[47] Ralf Schiffler, Quiver Representations, CMS Books in Mathematics (Ouvrages de Mathématiques de la SMC), Springer, Cham, 2014.

[48] N. Seiberg, Electric-magnetic duality in supersymmetric non-abelian gauge theories, Nucl. Phys. B 435 (1–2) (Feb 1995) 129–146.

[49] Rak-Kyeong Seong, Unsupervised machine learning techniques for exploring tropical coamoeba, brane tilings and Seiberg duality, Phys. Rev. D 108 (10) (2023) 106009.

[50] W.A. Stein, et al., Sage Mathematics Software (Version 9.4.0), The Sage Development Team, 2021, <http://www.sagemath.org>.

[51] Matthias Warkentin, Exchange Graphs via Quiver Mutation, Doctoral dissertation. Chemnitz University of Technology, Chemnitz, 2014.