

# Learning 3-manifold triangulations

Francesco Costantino<sup>1</sup>, Yang-Hui He<sup>2,3,4,5</sup> , Elli Heyes<sup>2,3</sup>   
and Edward Hirst<sup>6,\*</sup> 

<sup>1</sup> Institut de Mathématiques de Toulouse, 118 route de Narbonne, F-31062  
Toulouse, France

<sup>2</sup> London Institute for Mathematical Sciences, Royal Institution, London W1S 4BS,  
United Kingdom

<sup>3</sup> Department of Mathematics, City University of London, London EC1V 0HB,  
United Kingdom

<sup>4</sup> Merton College, University of Oxford, Oxford OX14JD, United Kingdom

<sup>5</sup> School of Physics, NanKai University, Tianjin 300071, People's Republic of China

<sup>6</sup> Centre for Theoretical Physics, Queen Mary University of London, London E1  
4NS, United Kingdom

E-mail: [e.hirst@qmul.ac.uk](mailto:e.hirst@qmul.ac.uk), [francesco.costantino@math.univ-toulouse.fr](mailto:francesco.costantino@math.univ-toulouse.fr),  
[hey@maths.ox.ac.uk](mailto:hey@maths.ox.ac.uk) and [elli.heyes@city.ac.uk](mailto:elli.heyes@city.ac.uk)

Received 24 July 2024; revised 16 January 2025

Accepted for publication 13 February 2025

Published 27 February 2025



CrossMark

## Abstract

Real 3-manifold triangulations can be uniquely represented by isomorphism signatures. Databases of these isomorphism signatures are generated for a variety of 3-manifolds and knot complements, using SnapPy and Regina, then these language-like inputs are used to train various machine learning architectures to differentiate the manifolds, as well as their Dehn surgeries, via their triangulations. Gradient saliency analysis then extracts key parts of this language-like encoding scheme from the trained models. The isomorphism signature databases are taken from the 3-manifolds' Pachner graphs, which are also generated in bulk for some selected manifolds of focus and for the subset of the SnapPy orientable cusped census with  $< 8$  initial tetrahedra. These Pachner graphs are further analysed through the lens of network science to identify new structure in the triangulation representation; in particular for the

\* Author to whom any correspondence should be addressed.



Original Content from this work may be used under the terms of the [Creative Commons Attribution 4.0 licence](https://creativecommons.org/licenses/by/4.0/). Any further distribution of this work must maintain attribution to the author(s) and the title of the work, journal citation and DOI.

hyperbolic case, a relation between the length of the shortest geodesic (systole) and the size of the Pachner graph's ball is observed.

Keywords: triangulations, machine learning, network science, 3-manifolds, isomorphism signature

## Contents

|   |    |
|---|----|
| 1. Introduction                                       | 2  |
| 2. Preliminaries                                      | 4  |
| 2.1. Triangulation IsoSigs                            | 4  |
| 2.2. Existing results surrounding Pachner graph paths | 6  |
| 3. Data generation & analysis                         | 7  |
| 3.1. Pachner graph network analysis                   | 9  |
| 3.1.1. Orientable cusped census                       | 14 |
| 4. ML IsoSigs   | 21 |
| 4.1. Differentiating manifolds                        | 22 |
| 4.1.1. NNs  | 24 |
| 4.1.2. Transformers                                   | 27 |
| 4.2. Differentiating Knots                            | 29 |
| 4.2.1. Supplementing with Dehn Surgery                | 30 |
| 5. Conclusions and outlook                            | 31 |
| Data availability statement                           | 32 |
| Acknowledgment  | 32 |
| Appendix A. IsoSig encoding                           | 32 |
| A.1. Example: triangular bipyramid                    | 34 |
| Appendix B. Pachner graphs                            | 35 |
| References  | 43 |

## 1. Introduction

Low-dimensional topology, as the study of manifolds of dimension 2, 3, and 4, has yielded countless invaluable results and insights. Initially it was driven by the quest of the Poincaré conjecture (stating that a 3-manifold homotopy equivalent to the 3-sphere is homeomorphic to it) and W. Thurston's geometrisation conjecture [1] which were finally proven by G. Perelman through Ricci flow regularisation techniques [2, 3]. But during the eighties the discovery of the first relevant examples of topological quantum field theories by Witten [4] established the field as a central testing ground for most physical theories. Therefore, beyond their intrinsic interest within mathematics, this field has long had large influence within physics [5], and more recently, its influence has been cemented in the development of the younger field of computer science [6], where in particular concepts such as triangulation and meshing are essential for rendering and visualisation.

As computational power has exponentially developed in recent decades, so too has the potential for implementation of resource-expensive computational methods. Many problems completely intractable before have now become rather accessible, largely through methods of 'big data' and 'big compute'. However mathematical research presents a unique challenge, where datasets can often be generated infinitely and sampled cheaply, exhaustive computation

becomes infeasible, demanding statistical practises. It is here, from a presently prominent sub-field of computational science that a wealth of new techniques have been made available for approaching mathematical research; that field being: machine learning (ML).

ML covers a broad range of techniques for computational statistics, and the subfield itself can be split into three core subsubfields, which each have already seen compelling successes in mathematics and mathematical physics. These are: supervised learning [7–26] (introduced in [27]), unsupervised learning [28, 29] (introduced in [30]), and reinforcement learning [31–36] (introduced in [37]); with notable introduction to string theory in [38–41]. This work focuses on techniques from the first of these subsubfields, in particular using supervised learning methods to classify and study 3-manifolds.

The direct application of ML methods within low-dimensional topology is increasingly widespread. Initiated in [42], where neural networks (NNs) were first used to predict knot invariants, after the surprisingly accurate prediction of the hyperbolic volume of a knot complement from its Jones polynomial [43], the study carried on in [44] where it was shown that NNs could effectively distinguish high complexity knot presentations of the trivial knot from non-trivial ones. In [45, 46], notable knot invariants (the Jones polynomial, Rasmussen’s  $s$ -invariant, and Khovanov homology) were mutually predicted by NNs with high accuracies, indicating new possible hidden relations between these invariants; and further knot invariant interrelations were identified in [47]. An especially useful application of these ideas has been developed in [48], where the recognition of ribbon knots has been pushed beyond the bounds of human capabilities through these techniques. Besides knots, ML methods have been used for the study of a particular kind of closed 3-manifolds [49], where graph 3-manifolds (3-manifolds whose Jaco–Shalen–Johansson decomposition contain no hyperbolic piece) were studied via their presentation through surgeries on plumbing graphs and various ML algorithms were implemented to reduce presentations of such 3-manifolds to easier ones via so-called Neumann’s moves.

In this paper a different approach to representation of 3-manifolds is considered, and one that has many direct practical applications: *triangulations*. As proved by Casler [50], every 3-manifold admits a triangulation, i.e. a decomposition into 3-dimensional simplexes (tetrahedra). By compactness only a finite number of these is sufficient and the (finite) datum of how to glue them can be summarised most succinctly in a single string of characters: the isomorphism signature (IsoSig). IsoSigs, defined in [51] and reexplained in appendix A, are especially compact encodings of 3-manifold triangulations, making bulk storage and study of triangulations far more accessible. Our first interest in this paper is to examine how amenable ML methods, and in particular NNs, are to extracting the highly-convoluted information about the triangulations from this highly-compressed representation.

Eight 3-manifolds were selected for focus in this study, chosen due to their importance within the field and variety of properties:  $S^3$ ,  $S^2 \times S^1$ ,  $\mathbb{RP}^3$ ,  $L(7, 1)$  and  $L(7, 2)$  (homotopy equivalent but not homeomorphic lens spaces), the three torus  $T^3$ , the Poincaré homology sphere PHS, and the ‘smallest closed hyperbolic’ 3-manifold with lowest volume  $H_{SC}$  (a.k.a. the Weeks manifold). After producing thousands of different triangulations of each of these manifolds we trained NNs (and transformers) to distinguish them pairwise, observing some striking performances, which gradient saliency techniques allowed interpretation of in terms of the IsoSig components (see section 4.1).

A single 3-manifold can be encoded via infinitely many distinct triangulations, where each IsoSig completely and uniquely encodes one of these triangulations. It is a classical theorem [52] that two such triangulations of the same 3-manifold can always be connected to each other by a finite sequence of local modifications, known as *Pachner moves*, which are of two types (the 2 – 3 move with its inverse, and the 1 – 4 move with its inverse). Therefore one can define

an abstract graph associated to each 3-manifold, known as the *Pachner graph*, whose nodes are triangulations and edges are associated to Pachner moves between them.

In this work, Pachner graphs were generated up to depth 5 for the eight 3-manifolds of focus, starting from a minimal triangulation. As well as these Pachner graphs providing the IsoSig data used in ML, their graph structure is of its own independent interest. Through a range of network analysis techniques these Pachner graphs were comparatively studied, with results in section 3.1. Techniques related to network clustering, shortest paths, centrality, and cycles were employed; notably identifying  $L(7, 2)$ 's most central triangulation was not the initial (minimal) one but one which was at distance two from it, containing 9 tetrahedra. Also, the Weeks manifold (and other hyperbolic manifolds not of focus here) exhibited significantly faster Pachner graph growth with depth, reaching computational limits and leading us to conjecture that Pachner graphs of closed hyperbolic 3-manifolds grow much faster than those of non-hyperbolic 3-manifolds.

Further investigations extended network analysis to the Pachner graphs of the snappy [53] census of cusped orientable hyperbolic 3-manifolds<sup>7</sup>, focusing on those with up to 7 tetrahedra (there are 4815 of them), see section 3.1.1. Our analyses evidenced an unexpected behavior which led us to conjecture 1 that the growth of the Pachner graph for these manifolds is inversely related to the length of the shortest closed geodesic (a.k.a. the systole). Additionally in section 4.2, knot complements were learnt via their cusped triangulation IsoSigs; focusing on eight knot complement 3-manifolds: the unknot, the trefoil knot, the figure eight knot, the knots  $5_1$ ,  $5_2$ ,  $6_1$ , and  $8_2$ . The results are comparable to those for closed 3-manifolds and interestingly show no learning in distinguishing the figure eight knot from the knot  $5_2$ , which are the two knots with smallest hyperbolic volume. We also supplemented the analysis for these knot complements by a similar analysis for their 0 and 1-surgeries: in section 4.2.1 we showed how the unknot and the trefoil knot are very accurately distinguished by these surgeries.

Computational work was completed in python, with use of the *regina* [54] and *snappy* [53] libraries for manipulating IsoSigs, as inspired by work in [51, 55]. Pachner graphs were constructed and analysed using *networkx* [56], whilst ML used the *sci-kit learn* [57] and *tensorflow* [58] libraries. Coding scripts and data are made available at this work's respective GitHub<sup>8</sup> repository, including an introductory notebook with useful functions for ones own further investigations manipulating and visualising Pachner graphs.

## 2. Preliminaries

### 2.1. Triangulation IsoSigs

Any compact connected 3-manifold  $M$  with boundary can be realised by gluing a finite set of  $T$  tetrahedra (3-simplices) along their faces via affine maps and then by deleting a small open neighborhood of the vertices of the tetrahedra. Such a structure is called an ideal triangulation and it induces a triangulation of  $\partial M$  by means of  $4T$  triangles (as each tetrahedron contributes 4 triangles to the triangulation of  $\partial M$ )<sup>9</sup>. In the special case of a closed manifold one can consider

<sup>7</sup> A word of caution here is needed: cusped hyperbolic 3-manifolds are encoded in *snappy* via their *ideal* triangulations, namely the (non-compact) manifolds are homeomorphic to the complement of the vertices in the triangulations encoded by the given IsoSigs. So our Pachner graphs describe such ideal triangulations for these cusped hyperbolic 3-manifolds.

<sup>8</sup> <https://github.com/edhirst/IsoSigPGML.git>.

<sup>9</sup> Since we allow self-gluing of tetrahedra, in some literature this is referred to as a ' $\Delta$ -triangulation'.

triangulations of  $M$  minus a finite set of disjoint open balls: indeed one can recover  $M$  from  $M \setminus B^3$  in a unique way by gluing back a  $B^3$  along its boundary  $S^2$ .

Each triangulation can be encoded by numbering the tetrahedra along with their vertices, indicating which pairs of faces of two tetrahedra of the triangulation are glued and how (there are 6 possible identifications for each pair of faces). All these data can be encoded in a single string of characters defined in [55], called the **IsoSig** of the triangulation.

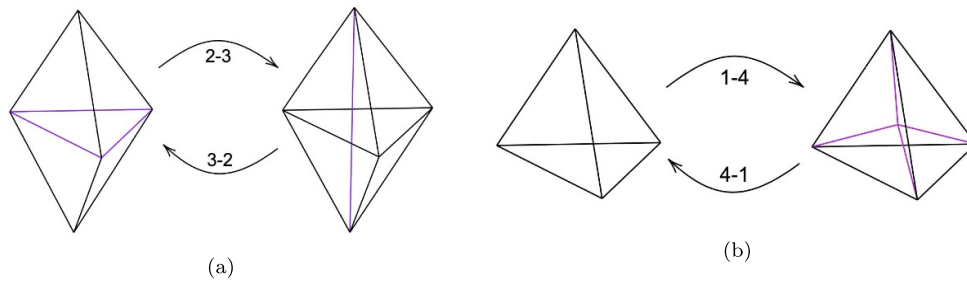
IsoSigs, are an exceptionally compact representation method for 3-manifold triangulations, introduced in detail with examples in appendix A, they are formed of a sequence of letters from the alphabet  $\{a, b, \dots, z, A, B, \dots, Z, 0, 1, \dots, 9, +, -, \}$ , encoding numbers in base 64 which enumerate in turn: the number of tetrahedra in the triangulation, the destination sequence of face gluings, the type sequence listing boundaries, and the permutation sequence dictating the vertex matchings of the face gluings. The encoding relies on many tricks for removing redundancies to make the IsoSig encoding especially efficient compared to traditional methods of listing each of these components separately. Importantly for the IsoSig representation, the following holds:

**Theorem (Theorem 6 in [55]).** *If two IsoSig strings  $S_1$  and  $S_2$  encode two triangulations then the triangulations are combinatorially isomorphic iff  $S_1 = S_2$ .*

Where two triangulations are isomorphic if they are combinatorially equivalent, such that some relabelling of tetrahedra, faces, edges, and vertices makes them the same. It should be noted though, that a 3-manifold admits infinitely many different (non-isomorphic) triangulations, which are all related to each other via the so-called bistellar or Pachner moves [52]. When applying a **Pachner** move, in general the IsoSig of a triangulation changes (for most moves even the number of tetrahedra changes). One can then build a graph  $\mathcal{P}$  whose nodes are the isomorphism types of triangulations of a fixed manifold and whose edges represent that the two triangulations can be transformed into one another via a possible Pachner move<sup>10</sup>. By the above cited fundamental result this graph is connected [52]. Before proceeding we remind the reader that these moves are of four types:  $2-3$ ,  $3-2$ ,  $1-4$  and  $4-1$  moves where for instance a  $2-3$  move consists in exchanging two tetrahedra with a suitable combination of three tetrahedra and a  $1-4$ -move consists in subdividing a tetrahedron into four tetrahedra by coning from its barycenter to its vertices. The moves  $3-2$  and  $4-1$  are the inverses of respectively the  $2-3$  and the  $1-4$ -moves. Note that the number of vertices of the triangulation does not change under  $2-3$  and  $3-2$  moves, but does change under  $1-4$  and  $4-1$  moves, exactly by one unit. Diagrammatics for these moves are shown in figure 1.

The ‘Pachner graph’  $\mathcal{P}$  described above is naturally endowed with two functions which we will denote  $v$  and  $t$  respectively given by the number of vertices and of tetrahedra of a triangulation. We will denote  $\mathcal{P}_n$  the full subgraph of  $\mathcal{P}$  of triangulations whose number of vertices is  $n$ : a particularly interesting subgraph is  $\mathcal{P}_1$ . At first sight one might be puzzled by this but we remind the reader that in our triangulations the tetrahedra are not embedded and self-identifications are possible. In fact it was proved independently by Matveev [59] and Piergallini [60] that for every 3-manifold the subgraph of  $\mathcal{P}_1$  formed by nodes corresponding to triangulations with at least 2 tetrahedra is non-empty and connected [59, 60]; there are only 3 triangulations with only 1 tetrahedra and they correspond to manifolds  $L(3, 1)$ ,  $L(5, 1)$  and  $S^3$ , so excluding them is not a problem.

<sup>10</sup> Note that there may be (likely rare) scenarios where two different moves can transform an input triangulation into the same output triangulation, in these cases we still only connect the respective nodes by a single edge.



**Figure 1.** Diagrammatics of the four Pachner moves, including (a) the 2 – 3 Pachner move with its inverse 3 – 2, and (b) the 1 – 4 Pachner move, with its inverse 4 – 1. For the 2 – 3 move the coloured lines bound a face which is replaced by an edge which then bounds the three tetrahedra inside. For the 1 – 4 move a vertex (coloured) is introduced into the centre and then connected to the remaining vertices, splitting the single tetrahedra into four.

## 2.2. Existing results surrounding Pachner graph paths

In this subsection we review some main results concerning the existence of algorithms for distinguishing 3-manifolds and their complexity. Our account is by no means complete nor exhaustive and we redirect the reader to the excellent survey by Lackenby [61] and to Matveev’s book [62]. The rough overall message is that although algorithms exist for basically all interesting tasks in 3-dimensional topology, when their theoretical complexity bounds are known, they are huge. Still, extensive experiments have been run for ‘small’ triangulations and they contrast sharply with the theoretical bounds.

One of the main facts to keep in mind is that a theoretical algorithm exists to tell whether two triangulations belong to the same 3-manifold or not. There are various proofs of this fact [63, 64], but they all use the proof of the Geometrisation Conjecture by Perelman. However, the best ‘known’ upper bound for the complexity of these algorithms, due to G. Kuperberg, is huge and is of the form  $2^{2^{\dots^t}}$  where  $t$  is the number of tetrahedra in the triangulations and the length of the tower of powers is a universal yet unknown constant [63].

The special case of recognising the 3-sphere is known to be NP and has a more explicit algorithm, due to Rubinstein [65]. The number of Pachner moves needed to connect a  $n$ -tetrahedra triangulation of  $S^3$  to the standard one has been bound by Mijatović [66] to be  $610^6 n^2 2^{20000n^2}$ .

The fact that an algorithm exists to tell whether two triangulations belong to the same manifold implies that there is also a theoretical computable function which bounds the distance of two triangulations in the Pachner graph of a given 3-manifold (see [61] theorem 4.3) based only on their numbers of tetrahedra. But for general 3-manifolds this is purely theoretical at present<sup>11</sup>.

We also mention that, although the theoretical bounds, when they exist, are huge, concrete experiments have been carried out and show quite a different behavior. In [55] Burton showed that if one restricts to 1-vertex triangulations with less than 10 tetrahedra (there are about  $8210^6$  of them!) then these can be connected to simpler ones (i.e. triangulations with less tetrahedra) by a sequence of Pachner moves whose length does not exceed 17 (of course this

<sup>11</sup> All the previous results can be largely improved when the 3-manifolds are link exteriors in the sphere, but we will not review the main results here as in this paper we are mainly concerned with closed orientable manifolds.

is possible only if one does not start from a minimal triangulation). In particular if one restricts to triangulations of  $S^3$  only 9 moves are needed. Furthermore in the same paper it is observed that the sequence of simplifying moves can be found so that the number of tetrahedra during the sequence never increases by more than 3.

Since a theoretical algorithm exists, but is likely exceptionally slow and computationally expensive, there is strong motivation for discovery of other methods for identifying equivalence of triangulations (that they represent the same 3-manifold and are in the same Pachner graph). Existing results show promise for identifying paths to triangulations with fewer tetrahedra, but not between triangulations with general numbers of tetrahedra. It is hence well motivated to tackle this broader general equivalence problem, where this work initiates the application of ML methods for this.

### 3. Data generation & analysis

To examine and learn the 3-manifold triangulation isomorphism via IsoSigs, we select a variety of 3-manifolds to focus on, in particular those which span a range of interesting properties. Eight 3-manifolds were selected for these investigations, they include: the 3-sphere  $S^3$ ; the product of the 2-sphere and the 1-sphere  $S^2 \times S^1$ ; 3-dimensional real projective space  $\mathbb{RP}^3$ , the lens space  $L(7, 1)$ , the lens space  $L(7, 2)$ ; the 3-torus  $T^3$ ; the Poincaré homology sphere PHS; and the smallest closed hyperbolic manifold  $H_{SC}$ . The 3-manifold abbreviations, along with the initial IsoSigs used to generate the datasets of equivalent triangulation IsoSigs are given in table 1. The IsoSig encoding is of particular interest due to its exceptionally compressed nature, and a thorough introduction to its structure is given in appendix A. Of these manifolds  $S^2 \times S^1$  is the only one not admitting a constant curvature isotropic complete Riemannian metric (we shall say for short a ‘non-isotropic manifold’), whilst the others are split amongst their curvatures, with  $(S^3, \mathbb{RP}^3, L(7, 1), L(7, 2), \text{PHS})$  positive curvature,  $T^3$  zero curvature, and  $H_{SC}$  negative curvature. In addition they have properties as listed in table 2.

This selection of manifolds is somewhat arbitrary, but includes a list of the prototypical manifolds often considered in other works. Specifically  $S^3$ ,  $T^3$ , represent the heuristically ‘simplest’ manifolds with  $\{+, 0\}$  curvatures, while  $H_{SC}$  is selected as a representative with  $-$  curvature. In fact, this choice is the hyperbolic 3-manifold with minimum volume (given in table 2, contrasting to the other selected manifolds which all have 0 volume).  $S^2 \times S^1$  is a common choice for considering composite manifolds which may be considered as a direct product of lower dimensional manifolds.  $\mathbb{RP}^3$  is the simplest 3d projective space, a particular important construction with a broad range of uses. Then, PHS is the only homology 3-sphere with finite fundamental group (other than  $S^3$  itself), making this an interesting comparison case for differentiating with  $S^3$ . Finally the two Lens spaces were the first constructions of 3-manifolds which were not differentiable by their homology and fundamental group alone (table 2 shows that these properties match). Under special (more trivial) conditions of the Lens space constructions both  $S^3$  and  $S^2 \times S^1$  can be produced, making these four a particularly interesting and potentially difficult set to differentiate between.

For each 3-manifold’s initial triangulation and IsoSig, a dataset was generated of equivalent triangulations (each represented with their respective IsoSig) from the manifold’s Pachner graph computed exhaustively up to as high a *depth* (i.e. integer number of moves away from the initial IsoSig) as feasible within memory limits. The Pachner graphs were generated with moves  $\{2-3, 3-2\}$  (with further functionality in the code to consider  $\{1-4, 4-1\}$  moves). Preliminary investigations by these authors for this work using supervised NNs to classify 3-manifolds from IsoSig triangulations showed promising performances with accuracies  $> 0.99$



**Table 1.** The eight 3-manifolds of focus in this investigation, listed with the initial 1-vertex triangulation IsoSigs used to generate their Pachner graphs, and the respective number of tetrahedra in these initial triangulations.

| Manifold Name           | Initial IsoSig              | Number of Tetrahedra |
|-------------------------|-----------------------------|----------------------|
| $S^3$                   | cMcabbgqs                   | 2                    |
| $S^2 \times S^1$        | cMcabbjaaj                  | 2                    |
| $\mathbb{RP}^3$         | cMcabbgqw                   | 2                    |
| $L(7,1)$                | eLkbcddhhjkh                | 4                    |
| $L(7,2)$                | cMcabbjqw                   | 2                    |
| $T^3$                   | gvLQQedfedffrwarhh          | 6                    |
| PHS                     | fvPQcdecdekrsnrs            | 5                    |
| $H_{SC} = \text{Weeks}$ | jLvAzQQcfeghighiiuquanobwwr | 9                    |

**Table 2.** table of selected manifold properties. Properties include the sign of the curvature (note  $S^2 \times S^1$  is non-isotropic so curvature is not applicable); Euler characteristic, hyperbolic volume, homology group, and fundamental group (given as generators and relators).

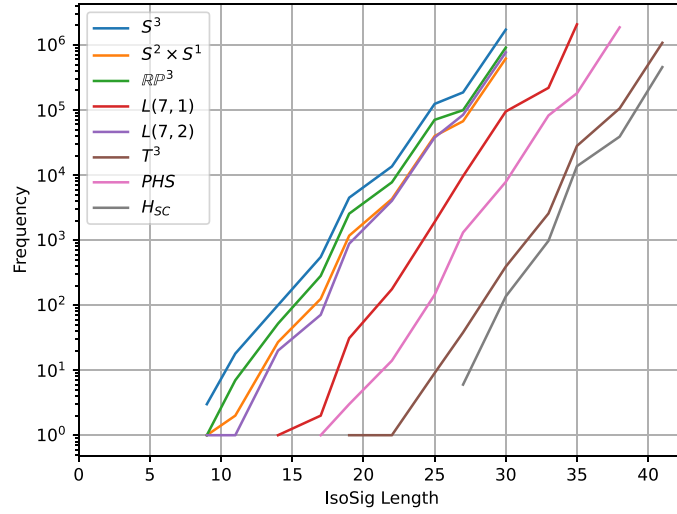
| Property   | Manifold |                  |                 |                |                |  | PHS     | $H_{SC}$                      |
|------------|----------|------------------|-----------------|----------------|----------------|--|---------|-------------------------------|
|            | $S^3$    | $S^2 \times S^1$ | $\mathbb{RP}^3$ | $L(7,1)$       | $L(7,2)$       | $T^3$                                  |         |                               |
| Curvature  | +        | ·                | +               | +              | +              | 0                                      | +       | —                             |
| Volume     | 0        | 0                | 0               | 0              | 0              | 0                                      | 0       | 0.94                          |
| Homology   | 0        | $\mathbb{Z}$     | $\mathbb{Z}/2$  | $\mathbb{Z}/7$ | $\mathbb{Z}/7$ | $\mathbb{Z} + \mathbb{Z} + \mathbb{Z}$ | 0       | $\mathbb{Z}/5 + \mathbb{Z}/5$ |
| Fund group | G        | x                | a               | a              | a              | a, b, c                                | a, b    | a, b                          |
|            | R        | x                | x               | aa             | aaaaaaa        | abAB,                                  | abbaB,  | ababAbbAb,                    |
|            |          |                  |                 |                |                | aCAc                                   | aaaabAb | abaBaaBab                     |

for near-perfect learning. However, these architectures relied heavily on the IsoSig length for distinguishing manifolds; where gradient saliency methods revealed that the only used input was the first padded entry indicating that change of IsoSig input length was the dominantly used feature in the classification.

Since the IsoSig length is heavily correlated with the number of tetrahedra, when we initiate generation using triangulations with a similar (e.g. small) number of tetrahedra then the graphs will display the same IsoSig lengths throughout, however the frequency distribution of the observed lengths varies, and this is what the architectures can leverage to provide efficient estimates of the considered manifold (note the final depth often all have the same length and dominate these distributions). For the manifolds considered the IsoSig lengths observed in the deep Pachner graph generation were: (9, 11, 14, 17, 19, 22, 25, 27, 30, 33, 35, 38, 41). The respective distributions are shown in figure 2.

Thus the ML focus was shifted to IsoSigs of the same length, such that the architectures could not shortcut to good performance and the true IsoSig language representation would have to be learnt. The datasets were created by randomly sampling all those of length 30 characters (the most evenly populous across the Pachner graphs) from the generated Pachner graphs (often exhibiting millions of IsoSigs). The sampled datasets of length 30 IsoSigs were all of size 2000, except for  $T^3$  with 391 and  $H_{SC}$  with 136, where this was the number of length 30 IsoSigs in the generated Pachner graphs. Therefore investigations with these two manifolds





**Figure 2.** IsoSig length frequency distributions across the deep Pachner graphs generated for the 8 manifolds considered.

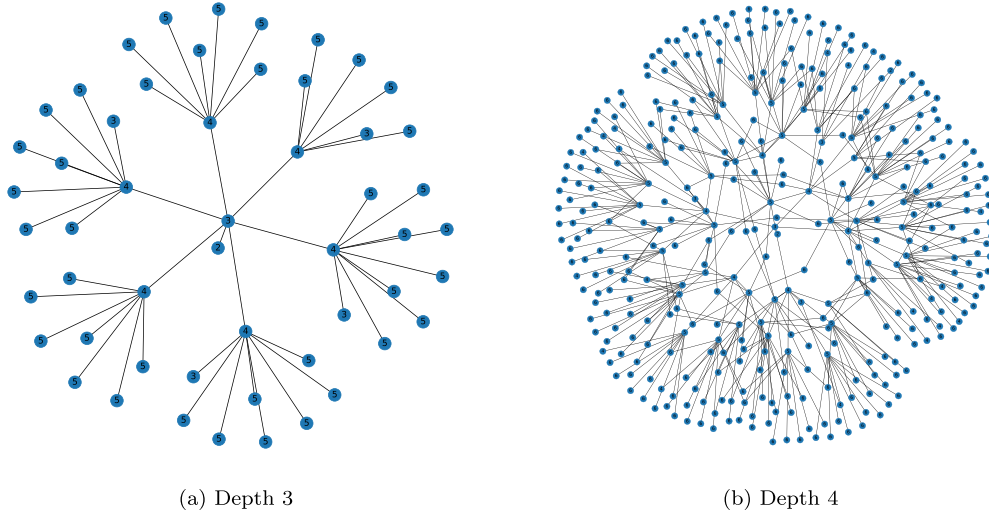
had bias in their datasets from these different sizes, which is mitigated by considering the MCC performance measure (over accuracy) which accounts for imbalanced classes [67].

The code scripts were written in python, using the *regina* library [54] for manipulating triangulations and applying Pachner moves, whilst also using the orientable cusped census hosted on *snappy* [53] to source some manifold IsoSigs. Scripts created Pachner graphs, as *networkx* objects [56], by exhaustively searching all vertices, edges, faces, and tetrahedra of each triangulation that Pachner moves could be applied to. Where performing a move generated a new triangulation the respective IsoSig was listed as a new node in the graph, connected by an edge (labelled with the move type) to the IsoSig the move was applied to. Conversely, where performing a move lead to a previously generated IsoSig an edge was introduced connecting the respective nodes already in the graph (ignoring the trivial inversion of moves). This Pachner graph generator function started from an initial input IsoSig and was adaptable to perform (input-specified) Pachner *moves* up to an input-specified *depth* from the initial IsoSig; outputting the Pachner graph and the list of IsoSigs, one for each triangulation of the manifold.

### 3.1. Pachner graph network analysis

The Pachner graphs used for IsoSig data generation are interesting combinatorial objects in their own right. As one performs Pachner moves with the hope of simplifying a triangulation, or looking to seek some specific triangulation property, one is effectively performing a walk on this graph. The graph network structure hence determines the potential success of these search algorithms, and more information about their network properties can guide the refinement of these search algorithms to make them as optimal as possible.

Therefore, with this in mind, we study these 1-vertex Pachner graphs  $\mathcal{P}_1$  for the considered 3-manifolds, using only  $\{2-3, 3-2\}$  moves, generated to as large a depth as can be performed in 240 core hours on a high-performance compute cluster [68]. Then with these large 1-vertex Pachner graphs we perform extensive analysis of their network properties: degrees, clustering, shortest paths, centrality, cycles. The distribution of number of tetrahedra in the



**Figure 3.** 1-vertex Pachner graphs for the  $S^3$  manifold, generated with  $\{2-3, 3-2\}$  moves up to the respective specified depths, from the initial triangulation with IsoSig: cMcabbgqs. The nodes are labelled with the number of tetrahedra in the respective triangulation.

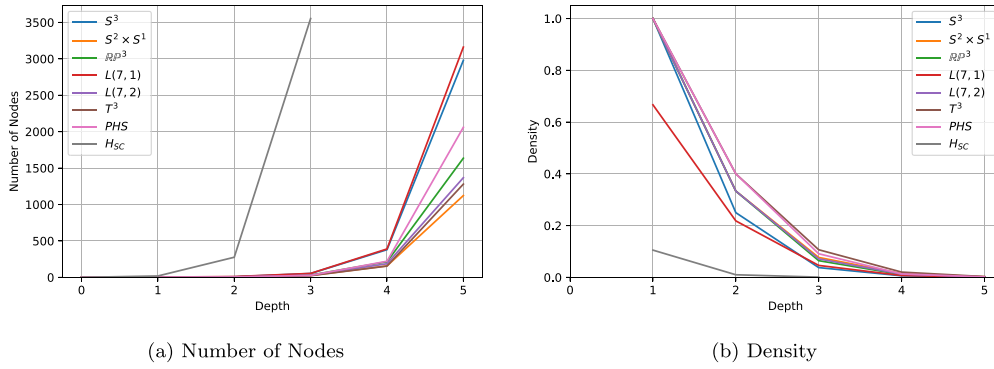
triangulations was also tracked, since seeking a 3-manifold triangulation with the minimum number of tetrahedra is highly sought-after in the field (for simplicity in computing topological properties and constructing composite 3-manifolds).

Example Pachner graphs for the  $S^3$  manifold, generated to depths 3 and 4 are shown in figure 3. The equivalent Pachner graphs for the remaining 7 considered manifolds are shown in appendix B. Each of these Pachner graphs were generated from the initial IsoSig given in table 1, using moves  $\{2-3, 3-2\}$ ; within these each Pachner graph node is an independent IsoSig / triangulation, with edges representing the moves connecting them (undirected since each  $3-2$  move changes the IsoSig in the opposite direction to the equivalent  $2-3$  move). The Pachner graph nodes are labelled with the number of tetrahedra in the respective triangulation. Since each of the initial IsoSigs represent 1-vertex triangulations, and only  $\{2-3, 3-2\}$  moves are considered which preserves the number of vertices in the triangulation, these graphs are (part of) the respective unique 1-vertex restricted Pachner graph  $\mathcal{P}_1$  for each manifold [55].

These Pachner graph images highlight well the rapid growth rate of triangulations. As can be seen for  $S^3$  in figure 3, it is only when depth 4 is considered that cycles begin to occur in this Pachner graph, at which point many cycles start to occur. These cycles also indicate shorter paths between different triangulations, compared to returning to the initial triangulation. The graphs also exemplify how number of tetrahedra tends to increase as one moves further from the initial triangulation in a Pachner graph.

To probe further the graph structure we turn to tools from network science, in particular applying an analysis technique from each of its core subfields: node analysis, clustering analysis, shortest path analysis, centrality analysis, cycle analysis. In addition, keeping track of the number of tetrahedra in each triangulation.

The Pachner graphs for the 8 selected manifolds of focus in this work were generated until memory limits deemed network analysis computationally infeasible (in each case these were



**Figure 4.** The growth rates of the Pachner graphs with depth for the 8 selected manifolds of focus in this work. Plot (a) shows how number of nodes increases with depth, whilst (b) shows the respective decreases in graph density.

RAM limits of 200GB). Fortunately many of these Pachner graphs failed generation at the same depth of 5, making comparison between them at this depth more meaningful. However, the final hyperbolic manifold  $H_{SC}$  grew too quickly and thus failed generation earlier, this makes its comparison less useful, but it is still included for completeness.

The Pachner rapid growth rate is well represented in figure 4, where the number of nodes in the Pachner graphs at each depth are shown. This plot clearly shows the final hyperbolic manifold having a much quicker growth rate leading to memory failure at lower depth. Generally, there will be more positions in a triangulation to apply a  $2-3$  move than a  $3-2$  move and so the number of tetrahedra tends to increase with depth, equivalently with more tetrahedra there are more positions to apply moves and thus this positive growth rate of triangulations in the Pachner graph appears. Respectively the graph density drops with depth, as each time a new node is introduced to the Pachner graph with  $N$  nodes,  $(N-1)$  potential edges are not introduced. As depth increases this contribution to density clearly beats any increase in density from introducing cycles, as displayed in the apparent convergence in density towards 0 with depth in figure 4.

These network analysis results are presented for the eight manifolds' 1-vertex Pachner graphs in table 3. This table lists the manifolds as well as the depth they were generated to (using input initial IsoSigs from table 1); each measure is discussed below with definitions.

**Nodes:** For each Pachner graph, the *number of nodes* is then listed, representing the number of unique triangulations generated. One striking fact is that the number of nodes for the depth 3 Pachner graph of the hyperbolic manifold  $H_{SC}$  is bigger than the number of nodes of the depth 5 Pachner graphs for the remaining (non-hyperbolic) manifolds. But for these remaining manifolds, we note the number of nodes does not seem to correlate with topological properties as the more populous graphs represent manifolds with varying curvatures, fundamental group sizes, etc; and the high number of  $S^3$  nodes contradicts an argument that it is based on the number of tetrahedra in the initial IsoSig. Respectively the graph *density* is listed, where density is the ratio between the number of edges in the graph and the number of edges in the complete graph with the same number of nodes (i.e. the maximum number of possible edges), which expresses the usual approximate inverse correlation behaviour with the number of nodes.

**Table 3.** Network analysis properties for the 1-vertex Pachner graphs of the 8 manifolds of focus in this work, generated to as a high depth as computationally feasible. Note that all were to depth 5, except for the last hyperbolic manifold which terminated generation at a lower depth. Analysis includes counts of the number of graph nodes, the graph density, the triangle and square clustering coefficients, the Wiener index in full and normalised form, the index of the most central node via eigenvector centrality as well as the range in centrality values across the graph, the decomposition of a minimum cycle basis, and the index of the node with the minimum number of tetrahedra and that minimum number.

| Manifold (depth)     | Number of nodes | Density | Clustering<br>(tri, squ) | Wiener index<br>(full, norm) | Centrality<br>(centre, range) | Minimum cycle basis<br>[[length, frequency]] | # Tetrahedra<br>(index, min) |
|----------------------|-----------------|---------|--------------------------|------------------------------|-------------------------------|--|------------------------------|
| $S^3$ (5)            | 2979            | 0.0009  | (0, 0.018)               | (30 734 710, 6.93)           | (7, 0.160)                    | [[4, 861], [6, 201], [8, 18]]                | (0, 2)                       |
| $S^2 \times S^1$ (5) | 1123            | 0.0023  | (0, 0.018)               | (4041 442, 6.41)             | (11, 0.220)                   | [[4, 305], [6, 61], [8, 2]]                  | (0, 2)                       |
| $\mathbb{RP}^3$ (5)  | 1636            | 0.0017  | (0, 0.018)               | (8738 384, 6.53)             | (13, 0.198)                   | [[4, 480], [6, 108], [8, 12]]                | (0, 2)                       |
| $L(7, 1)$ (5)        | 3161            | 0.0010  | (0, 0.022)               | (33 280 736, 6.66)           | (20, 0.201)                   | [[4, 1812], [6, 132], [8, 9]]                | (0, 4)                       |
| $L(7, 2)$ (5)        | 1368            | 0.0020  | (0, 0.019)               | (5996 840, 6.41)             | (2, 0.202)                    | [[4, 417], [6, 78], [8, 9]]                  | (0, 2)                       |
| $T^3$ (5)            | 1280            | 0.0030  | (0, 0.027)               | (4250 674, 5.19)             | (13, 0.215)                   | [[4, 1119], [6, 22]]                         | (0, 6)                       |
| PHS (5)              | 2060            | 0.0017  | (0, 0.022)               | (12 064 060, 5.69)           | (6, 0.215)                    | [[4, 1469], [6, 49]]                         | (0, 5)                       |
| $H_{SC}$ (3)         | 3553            | 0.0010  | (0, 0.016)               | (34 673 464, 5.49)           | (0, 0.230)                    | [[4, 2489], [6, 49]]                         | (0, 9)                       |

**Clustering:** Next the graph triangle and square *clustering coefficients* are listed [69, 70], which represent the proportion of 3 and 4 cycles that could exist in the graph that are there; i.e. all combinations of 3 nodes could be connected into a 3-cycle, the triangle clustering coefficient is the ratio between the number of 3-cycles in the graph to the number of combinations of choosing 3 nodes. The first observation is that all the Pachner graphs exhibit no 3-cycles, this is because any Pachner move changes the number of tetrahedra by 1, hence any 2 moves either change the number of tetrahedra by 2 or keep it the same (adding then removing, or removing then adding). Therefore any third move following these two must change the number of tetrahedra by 1 again, such that the total change in the number of tetrahedra after 3 moves cannot be 0 and then the same as the original triangulation. This argument actually lifts quite nicely to any odd cycle, such that all cycles in a 1-vertex Pachner graph must be even. The square clustering coefficients may then be non-zero, with all graphs exhibiting 4-cycles, and  $T^3$  having the highest proportion.

**Shortest Paths:** The *Wiener index* [71] is the sum of the lengths of the shortest paths over all pairs of nodes<sup>12</sup>, and is hence also presented in a normalised form with respect to the number of pairs of nodes (number of nodes  $N$  choose 2). A smaller normalised Wiener index indicates the graph is easier to walk around, with a high connectivity, which is the case for the non-positive curvature manifolds with the lowest values.

**Centrality:** Centrality measures determine the most central nodes, and how skewed the connectivity is away from uniform constant degree. The centrality measure examined in this work is *eigenvector centrality*, which assigns scores to the nodes based on their respective entries in the normalised eigenvector corresponding to the unique largest eigenvalue of the graph's adjacency matrix (via the Peron–Frobenius Theorem [72, 73]) [74]. A graph's adjacency matrix is intrinsically linked to potential paths through it, and such raising it to the  $k$ th power counts the number of  $k$ -paths between the indexed nodes. This limiting behaviour is then dictated by the largest eigenmode, and hence the eigenvector centrality score provides a means of determining the limiting connectivity of nodes. In table 3, the node index with the highest centrality score is listed, as well as the range of these scores across the graph. Interestingly, only for  $H_{SC}$  is the most central node the initial IsoSig triangulation, indicating the other Pachner graphs may be more symmetrically generated about alternative initial IsoSigs. In some cases, such as for  $L(7, 1)$  the index is as high as 20, such that it is the 20th triangulation produced when exhaustively generating the Pachner graph, and is 2 moves away from the initial triangulation; it is a 9-tetrahedra triangulation. Iteratively, this measure may be used in this way to search for useful starting triangulations. The range of centrality scores are largest for the non-positive curvature manifolds, which then turn out to have more skewed Pachner graphs and knowledge of the most central node becomes important as a key triangulation many paths will pass through (this could be a useful information when designing paths to minimise the number of tetrahedra).

**Cycles:** Cycles in a Pachner graph represent availability of different paths to move between selected nodes. As mentioned previously, since  $3 - 2$  and  $2 - 3$  moves each change the number of tetrahedra by 1, this leads to all cycles being even. 2-cycles enact the inverse of any move, so are trivial on these undirected graphs. This causes 4-cycles to be the first cycles that can occur,

<sup>12</sup> i.e. select a pair of nodes in the graph, find the shortest path between them (if there are multiple select one arbitrarily), count the number of edges in the path to get the length of this path, repeat for all pair of nodes in the graph, take the sum.

whose frequency is assessed through the square clustering coefficient. In general, a  $(2k)$ -cycle amounts to  $k$   $2-3$  moves and  $k$   $3-2$  moves and there are no *a priori* restrictions on their order. Where the number of moves is small relative to the number of tetrahedra, cycles are likely caused by the more trivial case where each of the  $k$   $2-3$  moves may be paired with its respective inverse  $3-2$  move removing the same tetrahedra introduced, but in a different order such that the intermediary triangulations are distinct. However as the number of moves gets larger there're less untouched tetrahedra to perform the new moves on, and then moves start to overlap in the same part of the triangulation leading to less trivial cycles.

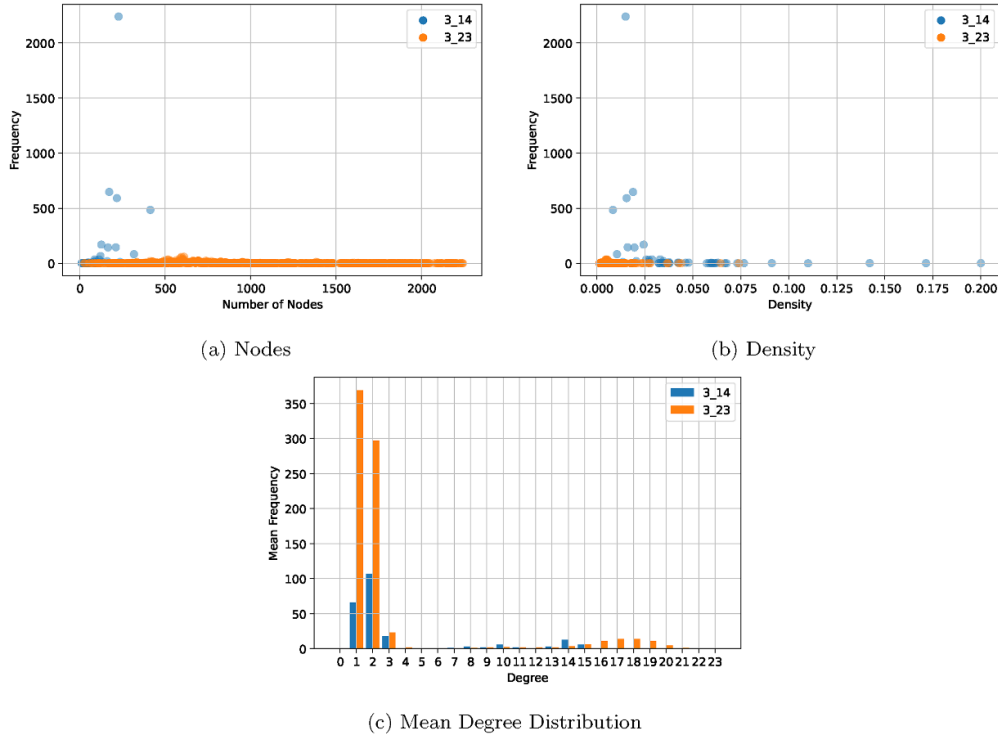
As can be seen in the Pachner graph plots, cycles often intersect, and where two smaller cycles intersect a larger cycle can easily be created by traversing the exterior of them both. To circumvent redundancies due to this, it is standard practice to consider a basis for the cycles [75], and additionally to focus on bases where the sum of the cycle lengths is minimised: a *minimum cycle basis*. However there is still redundancy in which cycles form part of this bases, and therefore the analysis in table 3 instead focuses on its properties, notably counting the frequencies of cycles of each length in the basis (which is unique across all minimum cycle bases). As can be seen, all the graphs exhibit 4-cycles, some with an order of magnitude more than others. Where the number of 4-cycles is higher the number of 6-cycles tends to be lower (except for  $\mathbb{RP}^3$  which has a high number of 6-cycles, and  $S^2 \times S^1$  with a surprisingly low number). Up to depth 5 the maximum cycle size which can occur is 10, and in none of the graphs does this occur, and interestingly it is only for the manifolds with simplest fundamental group presentation that 8-cycles occur.

**Number of Tetrahedra:** The final analysis tracks the node index with the fewest number of tetrahedra, often desirable for use in manifold computation and construction, as well as listing that number of tetrahedra. All manifolds have index 0 as their minimum (i.e. index 0 is the first triangulation in the generation process which is the initial triangulation), expected since a minimal triangulation was selected as an initial point for the generation. A final remark to emphasise is that the node with the minimum number of tetrahedra does not correlate with centrality, only in the case of  $H_{SC}$  does the most central triangulation have the minimum number of tetrahedra.

**3.1.1. Orientable cusped census.** Beyond the eight selected 3-manifolds that we focus on in this paper, the snappy Orientable Closed Census provides a list of 11 031 orientable closed hyperbolic manifolds. These however are represented as cusped manifolds which when filled have a range of tetrahedra for their initial triangulations from 9 to 37 – too large for sensible Pachner graph analysis as the number of triangulations would grow too quickly with depth to compute to a meaningful size.

Therefore we instead consider the respective snappy Orientable Cusped Census, which lists all 61 911 orientable cusped hyperbolic manifolds which can be represented by triangulations with at most 9 tetrahedra. We partition off the 4815 with  $< 8$  tetrahedra for an exhaustive analysis of their equivalent Pachner graphs. Each of these manifolds are represented by an initial triangulation (and hence IsoSig) in the database.

Despite the large number of manifolds in the census, with use of the high-performance compute cluster it became computationally feasible to generate Pachner graphs for *all* hyperbolic manifolds in this partition of the census, and perform the equivalent network analysis. Pachner graphs were generated for both move combinations  $\{2-3, 3-2\}$  and  $\{1-4, 4-1\}$ , up to depth 3, which we denote '3\_23' and '3\_14' respectively. The subsequent bulk network analysis provides insight into the triangulation deformation structure of these Pachner moves,

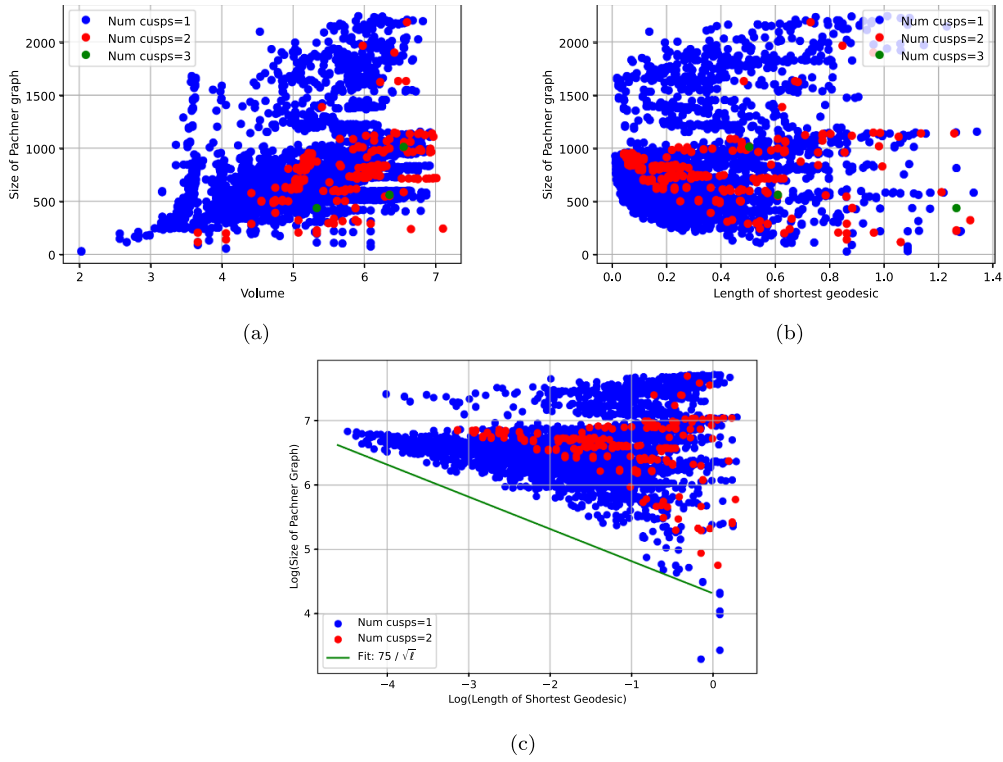


**Figure 5.** Network analysis of the orientable cusped census Pachner graphs (where number of initial tetrahedra  $< 8$ ). Analysis considers building the Pachner graphs to depth 3 with either only the  $\{1-4, 4-1\}$  moves (denoted 3\_14), or only the  $\{2-3, 3-2\}$  moves (denoted 3\_23) moves, plotting histograms of: (a) number of nodes in the graph; (b) graph density; and (c) the graph degree distribution averaged across the census. Note the degrees take integer values and the vertical lines are offsetted for visibility.

how they each impact the graph structure, as well as general insight into how Pachner graphs grow. Considering the same set of network analysis techniques used to examine the depth 5 Pachner graphs for the eight manifolds of focus in table 3, we created plots of these properties for the census as shown in figure 5.

**Nodes:** First order analysis of the graphs considers the sizes of the node and edge sets. A histogram for the number of nodes is shown in figure 5(a), demonstrating that the  $2-3$  (and their inverse  $3-2$ ) moves often lead to many more nodes. Since the number of possible  $1-4$  moves for a triangulation equals the number of tetrahedra  $N$ , whereas the number of possible  $2-3$  moves equals the number of common faces (which is half the total number of faces  $4N/2 = 2N$ ) this matches the expected approximate scaling that  $2N > N$  for triangulations of  $N$  tetrahedra, causing more possible  $2-3$  moves, and more nodes in the 3\_23 Pachner graphs. The  $1-4$  moves also have much higher frequencies, meaning their graph generation is more systematic, which is perhaps to be expected since the number of tetrahedra alone (which many manifolds will share) sets the number of  $1-4$  moves which can be enacted, whilst the specific triangulation gluing sets the number of  $2-3$  and  $3-2$  moves (and the rarer  $4-1$  moves). The density in figure 5(b) has a weak inverse relationship to the number of nodes, where adding nodes to a large graph introduces many more potential edges, and this respectively drops the

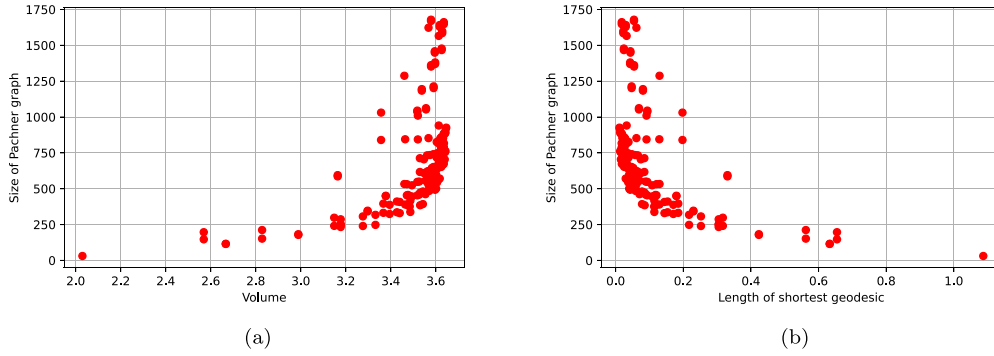




**Figure 6.** The distribution of the sizes of the Pachner graphs (i.e. number of nodes) as a function of (a) the volume and of (b) the length of the shortest geodesic (systole). In blue the points corresponding to one cusped manifolds, in red those corresponding to 2-cusped manifolds, and in green the remaining ones. Plot (c) shows plot (b) with axes logged to inform on the power law behaviour of the lower bound, including a potential lower bound of  $\frac{75}{\sqrt{\ell}}$ .

density value. Since the density does drop according to this behaviour this demonstrates that the Pachner graph edge sets grow slower than the node sets.

We also analysed the size of the Pachner graphs as a function of the hyperbolic volume of the manifolds and their number of cusps (which varies from 1 to 3 in our dataset). The result is shown in figure 6(a). Looking at the figure one realises that there are ‘peaks’ in the sizes of the Pachner graphs for one cusped manifolds and that these peaks correspond to the first red dots, namely the lowest volume 2-cusped manifolds. This is no random phenomenon. The blue dot with highest size of the Pachner graph and volume  $< 3.63$  turns out to be ‘v0117’ (in snappy’s nomenclature, IsoSig: hLALPkbcbefggghxwnxibk) and drilling out its shortest geodesic one gets ‘m129’ (IsoSig: eLPkbcddhggb) which is the second lowest volume manifold in snappy’s census with number of cusps 2 (it corresponds to one of the two leftmost red dots in the figure). This leads us to think that the size of the Pachner graph could be inversely related to the length of the shortest geodesic in the hyperbolic manifold which, by Thurston–Jorgensen’s theorem on hyperbolic Dehn fillings corresponds to Dehn fillings of manifolds with one more cusp with very large filling slopes. In order to check this we plotted in figure 6(b), the sizes of the Pachner graphs with respect to the length  $\ell$  of the shortest geodesic in the cusped manifold and the plot clearly shows that such number is bounded below by an



**Figure 7.** The distribution of the sizes of the Pachner graphs (i.e. number of nodes) as a function of (a) the volume and of (b) the length of the shortest geodesic restricted to the manifolds whose drilling along the shortest geodesic is a fixed one, here ‘m129’ (in *snappy*’s cusped census). The symmetry is not random and is related to the known relation between the volume of a manifold and its Dehn fillings along ‘long’ meridians.

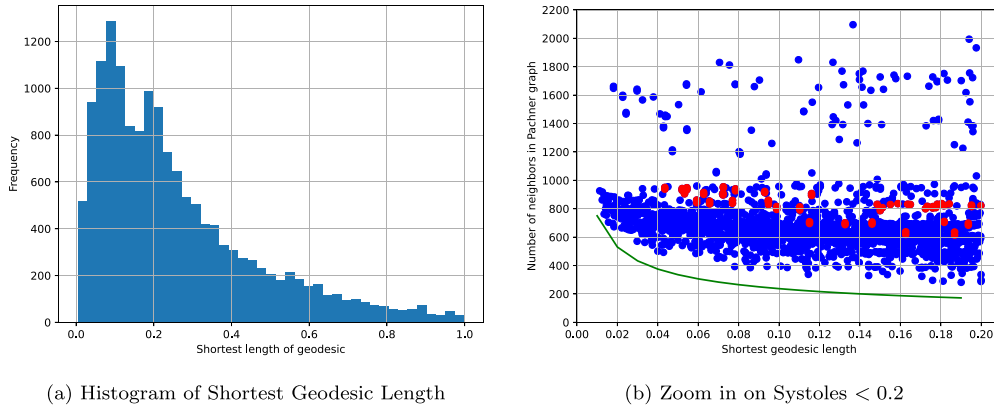
unknown decreasing function of  $\ell$ . Scaling the plot axes, as in figure 6(c), shows the behaviour more clearly; the lower bound forms a straight line indicating a power law relationship, and the gradient is  $\sim -0.5$ , indicating a relationship of the form  $\propto \frac{1}{\sqrt{\ell}}$  for this dataset.

Furthermore, a closer look reveals that the dots come in families, corresponding to which manifold is obtained by drilling the shortest geodesic. This is exemplified in figures 7(a) and (b) where the same analysis as before is restricted to the manifolds whose drilling is a fixed one, here ‘m129’ (chosen just as an example). All this seems to corroborate the following

**Conjecture 1.** The growth rate of the Pachner graph of a cusped hyperbolic 3-manifold  $M$  is bounded below by a function of the inverse of the systole (i.e. the length of the shortest geodesic in  $M$ ).

One may be concerned by a potential bias in the previous analysis: as the number of tetrahedra grows the relative scarcity of triangulations with big shortest geodesic length also increases, and this may be causing the observed phenomenon. To check this, we first drew the histogram of the shortest geodesic lengths for the manifolds with complexity up to 8 in the Orientable Cusped Census of *snappy*: see figure 8(a). The figure clearly shows that triangulations with shortest geodesic  $< 0.2$  are fairly common, thus in the right part of the same figure we restrict the image of figure 7(b) to the interval  $\ell \in [0, 0.2]$ . The results are consistent with the observed behaviour.

**Degrees:** The node degree distribution of a graph can be described by a histogram showing the frequency of the nodes that have a certain number of edges incident to them. Averaging the degree distributions over the entire considered census partition produces the plot given in figure 5(c) (i.e. the mean frequency of 3\_23 nodes with degree 1 across the 4815 manifolds in this census subset is 368.9). For both move styles all graphs had at least one edge, so that a move of either form could always be performed on the initial triangulation. This is shown by a 0 mean frequency of degree zero, so that no Pachner graphs were the trivial empty graph formed by a single node and no edges. For both move styles the degree distributions were bimodal, with 2 peaks, indicating that some moves (in both cases) significantly open up the number of new subsequent move possibilities and thus have much higher degree in this exhaustive graph generation. The peaks in the distributions (particularly for

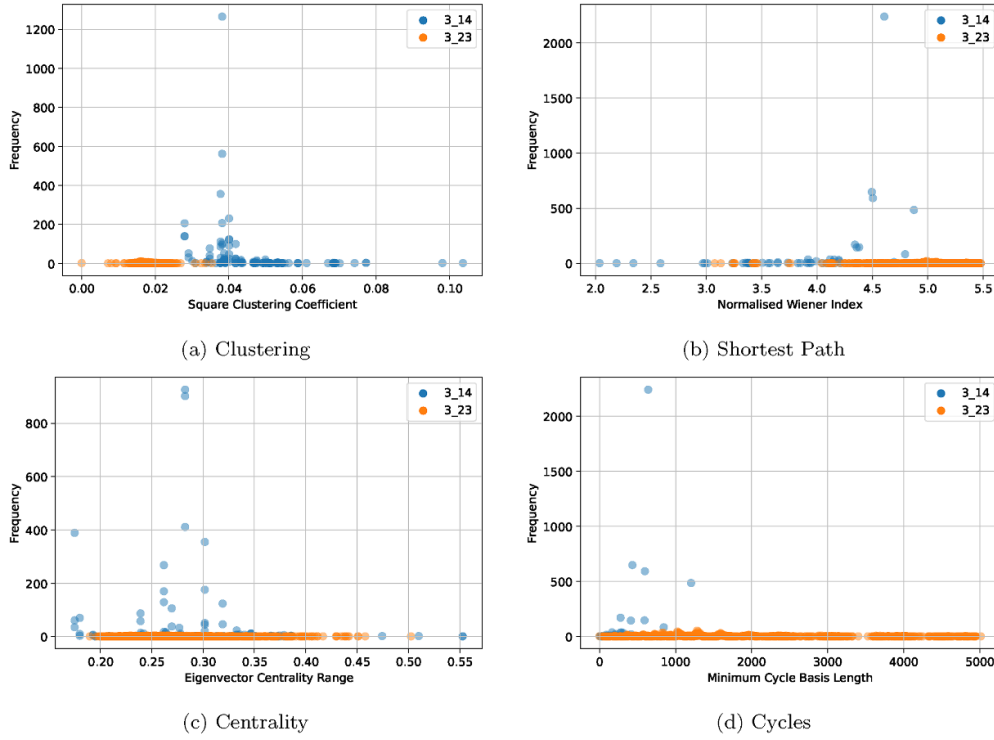


**Figure 8.** On the left the global histogram of the shortest geodesic length on the Oriented Cusped Census of snappy with  $< 9$  tetrahedra. On the right the restriction of the plot of the size of the Pachner graph for manifolds with shortest geodesic less than 0.2. In green, the graph of the function  $f(\ell) = \frac{75}{\sqrt{\ell}}$  seems to approximately minor the scatter plot on the interval  $[0, 0.2]$ , but it is plot here only as an example.

degree 1) are due to the leaves of the Pachner graphs created in the latter moves at the limit of the generation truncation, however they can occur otherwise as exemplified in figure 3(a). The 3<sub>23</sub> graphs have a larger relative size between the bimodal peaks than the 3<sub>14</sub> graphs ( $368.90/13.94 = 26.46 > 8.48 = 107.04/12.63$ ), so that the 3<sub>14</sub> graphs' nodes are more skewed to higher degrees as these moves are more likely to open up many move possibilities, despite there being less possibilities demonstrated by the degree value for the second peak. The 3<sub>23</sub> graphs also have a larger difference between their two peaks ( $18 - 1 > 14 - 2$ ), indicating that the edges are less evenly distributed between the nodes. This is linked to the move possibility scaling, where as a  $2 - 3$  move changes the number of tetrahedra in a triangulation from  $N$  to  $(N + 1)$  the approximate scaling of the  $2 - 3$  move possibilities grows at a rate of order  $2N$  (i.e. the number of faces mod the number of tetrahedra per face,  $4/2 = 2$ ). Alternatively as a  $1 - 4$  move changes the number of tetrahedra from  $N$  to  $N + 3$  the scaling of the  $1 - 4$  move possibilities grows exactly at a rate of  $N$ . Therefore, particularly as moves increase the number of tetrahedra, there are on average relatively more  $2 - 3$  moves which can be made, leading to higher degree nodes<sup>13</sup>.

**Clustering:** As in the previous Pachner graph analysis the triangle clustering coefficients for graphs across the census were all zero, since these graphs can only exhibit even length cycles. However the square clustering coefficients had (min,mean,max) across the census for the 3<sub>14</sub> graphs of (0.028, 0.038, 0.104) and for the 3<sub>23</sub> graphs of (0.000, 0.018, 0.035), with only the 3<sub>23</sub> graphs having examples of 0 square clustering coefficients, occurring for 2 manifolds. The full distributions are shown in figure 9(a), they show that 4-cycles are always introduced with  $1 - 4$  moves since all the initial triangulations have at least 2 tetrahedra so a  $1 - 4$  move on each followed by their inverses in the other order automatically creates a 4-cycle. In general 4-cycles are less likely with  $2 - 3$  moves since the gluings change more non-trivially and

<sup>13</sup> Note in this crude leading behaviour approximation we neglect contributions from the less probably inverse  $3 - 2$  and  $4 - 1$  moves.



**Figure 9.** Core network analysis properties of the Pachner graphs for the orientable cusped census partition with number of initial tetrahedra  $< 8$ . Analysis considers building the Pachner graphs to depth 3 with either the  $\{1 - 4, 4 - 1\}$  (denoted 3\_14) or the  $\{2 - 3, 3 - 2\}$  (denoted 3\_23) moves, plotting histograms of: (a) the square clustering coefficients; (b) the normalised Wiener index; (c) the range of eigenvector centrality values; and (d) the length of the minimum cycle basis.

hence alternating inverses may be obstructed. The 3\_14 graphs again have higher frequencies, indicating they are more likely to have similar graph structure.

**Shortest Paths:** The normalised Wiener index provides a global measure of the average shortest path length between any two nodes in the Pachner graph, that can be appropriately compared between graphs. The (min,mean,max) values for the 3\_14 graphs were (2.04,4.54,4.87) and for the 3\_23 graphs were (3.08,5.08,5.48). In figure 9(b) a histogram of these values are plotted also, showing that the 3\_14 graphs tend to have shorter shortest paths between nodes so that they can be traversed more easily, and it may be better to enact first these moves were possible in triangulation simplification methods. The 3\_14 graphs have higher frequencies, again supporting that graphs created by these moves are more similar and less dependent on the underlying manifold topology.

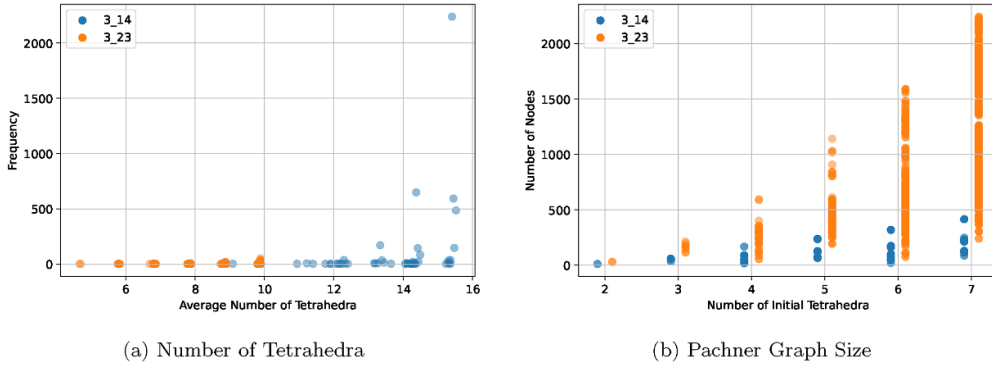
**Centrality:** Eigenvector centrality uses the eigenvector values corresponding to the adjacency matrix's unique largest eigenvalue to assign centrality scores. Since these scores are relative their absolute values are less relevant, and instead focus in this work is placed on the range of values across a given Pachner graph (highest – lowest centrality score). The (min,mean,max) values of the centrality ranges for the 3\_14 graphs were (0.175,0.271,0.553) and for the 3\_23

graphs were  $(0.190, 0.261, 0.502)$ ; with 487 of the 4815  $3_{14}$  graphs having the initial triangulation as the most central, and 485 times for the  $3_{23}$  graphs. Therefore, of the 4815 manifolds only  $\sim \frac{1}{4}$  have the initial node as the most central node in both cases, implying the generation process is highly non-symmetric. The similar centrality ranges for both move styles indicate similar adjacency eigendecompositions. Again higher  $3_{14}$  frequencies indicate similar graph structures.

**Cycles:** As previously observed, Pachner graphs can only exhibit even cycles, and additionally 2-cycles are not possible in this Pachner graph design where undirected edges represent the ability to move between triangulations (not a unique move). Since the graphs generated in the census were truncated at depth 3, the largest cycle size possible is 6, and hence the cycle bases for all these graphs were spanned by 4-cycles and 6-cycles. The average number of 4 and 6-cycles in each minimum cycle basis for the census partition's generated Pachner graphs were  $(155.16, 0.00)$  for  $3_{14}$  graphs and  $(354.58, 12.87)$  for  $3_{23}$  graphs. These results show that although the Pachner graphs are only generated to depth 3 they are highly interconnected with a huge number of cycles. The  $3_{23}$  graphs have more cycles on average as there are many more move choices, and hence much more path degeneracy. Surprisingly the  $3_{14}$  graphs never exhibit 6-cycles, presumably because in all scenarios they are visible they were decomposable into 4-cycles in the basis length minimisation process (as the reordered inversion of any  $3_{1-4}$  moves to make a 6-cycle would have all reorderings of 2 moves making sub 4-cycles). Higher frequencies for these  $3_{14}$  graphs again indicate similar graph structures with similar cycle decompositions, less dependent on underlying topology.

**Number of Tetrahedra:** The number of tetrahedra in each triangulation across a Pachner graph can only take 1 of 4 values (as only generated up to depth 3), based on an initial triangulation with  $N$  tetrahedra and a move set of  $\{2-3, 3-2\}$  then the values are  $\{N, N+1, N+2, N+3\}$ , alternatively for move set  $\{1-4, 4-1\}$  the values are  $\{N, N+3, N+6, N+9\}$ . This is because the moves always change the number of tetrahedra by a fixed amount, and therefore the number of tetrahedra in the initial triangulation is paramount to determining the spectrum, which is selected for the dataset to be a local minimum. We note that this is not always a unique minimum as the frequency of triangulations with as many tetrahedra as the initial triangulation is occasionally  $> 1$ . The number of tetrahedra in the initial triangulation takes values of  $(2, 3, 4, 5, 6, 7)$  across the census partition, with respective frequencies  $(2, 9, 56, 234, 962, 3552)$ . Therefore most manifolds are initialised with triangulations with 7 tetrahedra.

Figure 10(a) shows a histogram of the average number of tetrahedra across each Pachner graph, plotted for all members of the census partition. It shows that the  $3_{14}$  graphs have a higher average number of tetrahedra, expected since the  $1-4$  moves introduce more tetrahedra than the  $2-3$  moves (and both these are more common than their inverses). There are also higher frequencies for these graphs as the graph structures are similar with each tetrahedron available to perform a  $1-4$  move irrespective of its gluing. Moreover, figure 10(b) shows how the number of tetrahedra in the initial triangulation correlates with the number of nodes in the generated Pachner graph. Generally, more initial tetrahedra leads to more nodes and larger graphs, with larger ranges for the graph sizes as the initial number increases; this is related to more tetrahedra providing more possibilities for moves, facilitating the rapid graph growth with depth. Since the number of places to apply  $2-3$  moves grows quicker than for  $1-4$  moves (each new tetrahedra introduce  $> 1$  new common faces) these  $3_{23}$  graphs are more responsive to the initial number of tetrahedra but also exhibit a greater range of values as the combinatorics is more susceptible to limitations caused by complicated gluings.



**Figure 10.** Distributions of number of tetrahedra in triangulations across the 4815 Pachner graphs of the orientable cusped census partition with number of initial tetrahedra  $< 8$ . Analysis considers building the Pachner graphs to depth 3 with either the  $\{1-4, 4-1\}$  (denoted 3\_14) or the  $\{2-3, 3-2\}$  (denoted 3\_23) moves, plotting (a) a histogram of the average number of tetrahedra across each Pachner graph; and (b) the number of tetrahedra in the initial triangulation against the number of nodes in the Pachner graph. Note the number of initial tetrahedra take integer values and the vertical lines are offsetted for visibility.

Considering the limiting behaviour in the 3\_23 graphs of a tree-like Pachner graph (no IsoSigs are repeated during generation) for the case of only considering  $2-3$  moves (ignoring the  $3-2$  inverse move), if one were to start with a triangulation of  $N$  tetrahedra, in this limiting case this means there are  $2N$  common faces to perform  $2-3$  moves and hence  $2N$  nodes at depth 1 of the Pachner graph. Each of these triangulations will have  $N+1$  tetrahedra, and hence  $2(N+1)$  faces for  $2-3$  moves, leading to  $4N(N+1)$  nodes at depth 2. Similarly, these triangulations will now have  $N+2$  tetrahedra,  $2(N+2)$  faces for  $2-3$  moves, leading to  $8N(N+1)(N+2)$  nodes at depth 3. In this limiting case one would then expect  $8N(N+1)(N+2) + 4N(N+1) + 2N + 1 = 8N^3 + 28N^2 + 22N + 1 = 8N^3 + \mathcal{O}(N^2)$  nodes of the Pachner graph. Mapping this leading order behaviour to the initial  $N$  values considered in figure 10(b), the predicted number of nodes would respectively be (64, 216, 512, 1000, 1728, 2744) for the  $N$  range  $[2, 7]$ , which does approximately match the observed maximum counts in the plots. The deviations from the maxima are when this limiting tree-like Pachner graph approximation breaks down, as triangulations are repeated in generation and the graph exhibits cycles (as well as minor effects from the less frequently occurring  $3-2$  moves). Equivalently in the 3\_14 graphs for the limit case of only  $1-4$  moves in a tree-like Pachner graph, an initial triangulation with  $N$  tetrahedra has exactly this many tetrahedra to perform  $1-4$  moves, leading to  $N$  Pachner graph nodes at depth 1. Respectively there are then  $N(N+1)$  nodes at depth 2, and  $N(N+1)(N+2)$  at depth 3, for an overall estimate of  $N^3 + 4N^2 + 4N + 1 = N^3 + \mathcal{O}(N^2)$ . The respective estimates would then be (8, 27, 64, 125, 216, 343), again matching the observed order of the maxima in figure 10(b).

#### 4. ML IsoSigs

In recent decades, ML methods have seen a wealth of successes across many fields academia. This can largely be attributed to the exponential growth in computational power, allowing the generation of ever-larger datasets for broader and more thorough analysis.

Mathematical data has a unique advantage in that it is exact, there is no noise from experimental apparatus, and as such any statistical patterns in data can be interpreted to have more direct meaning about the data's underlying structure. However, new mathematical data can often be generated infinitely, and thus exhaustive computation becomes impossible, relying on appropriate sampling. Hence there is a true need for statistical methods when performing analysis on these large datasets. The computational statistical methods we explore here come under ML.

In this work we examine the IsoSig representation of 3-manifold triangulations, testing how well can ML methods learn to identify manifolds from these IsoSigs. Already, as demonstrated in section 2.2, identifying the equivalence of the triangulations is highly computationally intensive. Where sequences of Pachner moves between two triangulations are not known to exist, the results from ML can provide a probability of existence of such a sequence moves. Successes in learning IsoSig equivalence for 3-manifolds then opens up the scope for development into learning manifold properties and finding explicitly the connecting paths between triangulations.

For these investigations the IsoSig input data is paired with an output label which identifies to which manifold the triangulation belongs to. The ML problem is in the style of supervised classification, where the network seeks to sort IsoSigs corresponding to different manifolds into separate classes. In this setup, the networks are learning the combinatoric structure of triangulations which differentiates the manifolds, directly linked to their respective topology.

Specifically, for each of the eight 3-manifolds considered, a large Pachner graph was generated to depth 8 as described in section 3. From there the frequency distribution of IsoSig lengths was computed across all IsoSigs in each manifold's Pachner graph. Then an appropriate length was selected such that sufficiently large samples could be taken from each Pachner graph to produce datasets of fixed length IsoSigs<sup>14</sup>; the fixed length considered was 30 characters.

For computational processing these IsoSigs were converted to arrays of integers, using one-hot encoding of each character in the IsoSig. Since the vocabulary consists of 64 characters  $\{a, b, c, \dots, z, A, B, C, \dots, Z, 0, 1, 2, \dots, 9, +, -\}$ , the encoding of each character is a 64-dimensional sparse vector of 0s with a single 1 in the dimension indexing that character in the vocabulary. Therefore, because all IsoSigs considered for ML were 30 characters long, each IsoSig was encoded by a  $30 \times 64 = 1920$ -dimensional vector. Despite this representation being rather memory intensive, requiring more computational resources for training any architecture, the naturalness of the one-hot encoding style (where each character is an independent dimension) makes learning clearer and more interpretable<sup>15</sup>.

#### 4.1. Differentiating manifolds

In this section, the ML architectures of NNs and *transformers* are both applied to the problem of binary classifying IsoSigs belonging to different manifolds. For the eight 3-manifolds focused on in this work, this creates  $\binom{8}{2} = 28$  pairs of manifolds to design paired data for, and perform binary classification on. Creating 28 binary datasets to learn:  $\{\text{IsoSig}\} \mapsto \{0, 1\}$ , the performance can then be compared between different manifold pairings, providing insight

<sup>14</sup> As noted previously, initial testing showed that padding mixed length IsoSigs caused the architectures to focus exclusively on the distribution of lengths, which was sufficiently different between manifolds to lead to good classification results. This simplification was avoided by considering a fixed length.

<sup>15</sup> The other natural encoding of just mapping each character to a number  $0 \mapsto 63$  was tested however lead to lower learning performance scores.



into what topological properties about the manifolds, encoded in the IsoSigs, the networks use to differentiate between manifolds.

The NN architecture is the prototypical ML model, it is very adaptable, and used ubiquitously with great success throughout many fields. Any NN is made up of sets of neurons organised into layers, where each neuron takes in a vector of inputs, and outputs a single number. The neuron acts on the input vector  $\mathbf{x}$  with a linear function, then a non-linear ‘activation’ function, such that:  $\mathbf{x} \mapsto \text{act}(\mathbf{w} \cdot \mathbf{x} + b)$  for the trainable parameters of weights vector  $\mathbf{w}$  and bias scalar  $b$ . The full NN function of  $L$ -layers then takes the form

$$f_{\text{NN}} := \text{softmax} \left( W_{i_L, i_{L-1}} \cdot \alpha \left( \dots \alpha \left( W_{i_1, i_{\text{input}}} \cdot x_{\text{input}} + b_{i_1} \right) \dots \right) + b_{i_L} \right). \quad (4.1)$$

Demonstrating the vectorised linear action of matrix multiplication with the weight matrix  $W_{i_m i_n}$  between layers  $i_m$  and  $i_n$ , then addition of a bias vector  $b_{i_m}$ . Followed by the  $\alpha$  activation function used to provide the alternating non-linear action; here we use the Leaky-ReLU activation function, which has significant speed advantages due to its simplicity, and effectively makes the training process the fitting of a piecewise linear function. Finally the softmax activation function is used to bound outputs into the range  $[0, 1]$  such that they produce a probability distribution for the classification task. These activation functions are defined

$$\text{Leaky-ReLU}(x)_i = \begin{cases} x_i & x > 0 \\ \varepsilon x & \text{otherwise} \end{cases}, \quad (4.2)$$

$$\text{softmax}(x)_i = \frac{e^{x_i}}{\sum_j e^{x_j}}, \quad (4.3)$$

for some small factor  $\varepsilon$  ( $\sim 0.01$ ). These architectures are very adaptable, possess some universal approximation theorems supporting their use, and have practically been particularly successful across other fields of mathematics. More information on this architecture can be found at [76].

Conversely, the transformer architecture is much more sophisticated, and involves many stages of sub-NNs for encoding and decoding, primarily relying on the concept of attention [77] to learn context amongst the input strings—note transformers do not require the tensorial encoding of the IsoSigs, but do require substantial computational resources, so are tested less extensively in this work.

Generally, transformer models are another class of supervised architectures designed to handle sequential language-like data using attention mechanisms instead of traditional recurrence or convolution. At their core, they employ an encoder–decoder structure, though many applications focus on encoder-only (e.g. BERT) or decoder-only (e.g. GPT) variants. The multi-head self-attention mechanism is the cornerstone, which allows the model to dynamically weigh the importance of all input tokens relative to each other. For an input sequence  $abcd \dots$  attention scores are computed using query  $Q$ , key  $K$ , and value  $V$  matrices, from which the attention importance weighting between letters of the input is computed as

$$\text{Attention}(Q, K, V) := \text{softmax} \left( \frac{QK^T}{\sqrt{d_k}} \right) V, \quad (4.4)$$

for  $d_k$  the dimensionality of the key vectors. This mechanism enables Transformers to capture long-range dependencies efficiently.

Transformers work via alternating small NNs providing trainable parameters for the encoding and decoding, and attention layers which preserve the interrelations within the sequence. Transformers are trained using a supervised loss function in a similar manner to NNs, typically using cross-entropy loss, over a labelled dataset. They are also particularly popular as the

attention mechanism allows parallelisation in computation, making larger models more feasible than traditional recurrent methods. More information on this architecture can be found at [78].

As motivated in section 2.2, the task of identifying equivalent triangulations is infeasibly expensive, and rigorous algorithm design is impractical. Therefore efficient statistical methods may prove useful in informing on the likelihood of equivalence, thereby practically informing on optimal resource allocation for more rigorous checks; i.e. if a statistical method says a pair IsoSigs are very likely to be equivalent, it may be worth committing resources to compute the connecting path where it would be less likely found for other pairs. Recent years have seen ML dominate the development of computational statistical methods across fields, and in this section we provide the first investigation into its efficacy on this equivalence problem and for general 3-manifolds as a whole. In doing so we consider<sup>16</sup> the ubiquitous NN architecture, and the transformer architecture which has shown particular success on language-like data (alike our IsoSig encodings).

**4.1.1. NNs.** The NNs used in these investigations were built using `tensorflow` [58], with architecture of three dense layers (sizes 256, 128, 64) then a final dense layer of size 2 for the one-hot encoded binary output; all layers used leaky-ReLU activation (factor 0.01), L2 regularisation, and dropout (factor 0.01) – except for the output layer using softmax activation for the binary classification. The NNs were trained with an Adam optimiser to minimise the binary cross-entropy loss for 30 epochs with a batch size of 64 and a learning rate of 0.001.

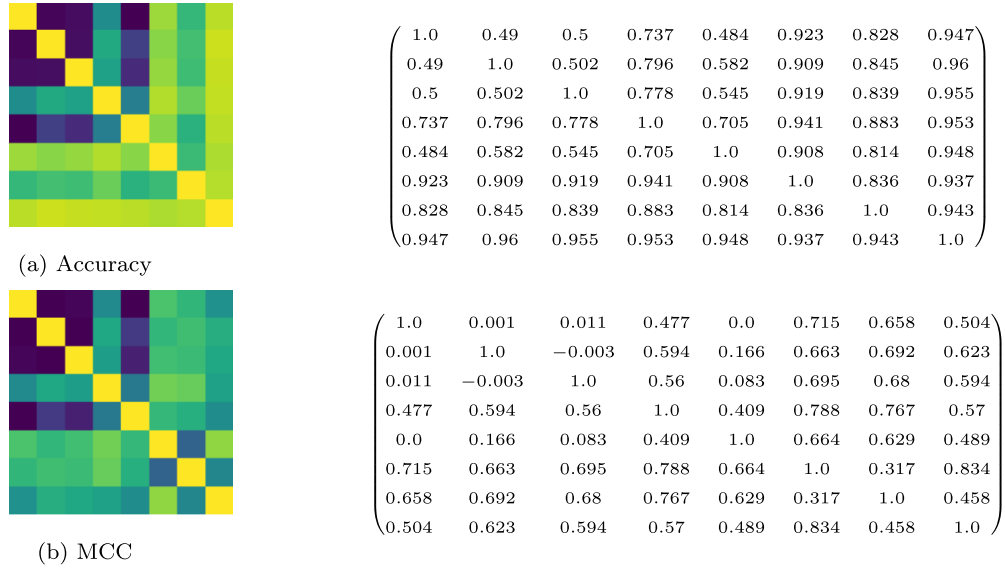
Binary classification performances were measured with accuracy (the proportion of correctly classified test IsoSigs), and Matthew’s correlation coefficient MCC (an unbiased alternative to accuracy). Both measures evaluate to 1 for perfect learning, whereas for no learning in a binary classification problem accuracy is 0.5 and MCC is 0. Each binary classification investigation was performed 5 times on 5 different partitions of the IsoSig data into train and test sets, providing statistical confidence in the mean performance measures reported (this process is known as cross-validation).

The performance measure results are arranged in a matrix format in figure 11. Each entry in the matrix  $M_{ij}$  represents the mean measure score across the cross-validation runs for the binary classification of IsoSigs pairs between manifolds with index  $i$  and index  $j$  in the list:  $(S^3, S^2 \times S^1, \mathbb{RP}^3, L(7, 1), L(7, 2), T^3, \text{PHS}, H_{\text{SC}})$ <sup>17</sup>. Heat maps of these matrices are also shown for ease of comparison.

For each matrix the off-diagonal scores respectively have (min, mean, max): (a) Accuracy (0.4835, 0.800, 0.960), (b) MCC (-0.003, 0.487, 0.834); showing a range from no learning to near-perfect learning. Comparing between manifolds, the positive curvature manifolds are harder to differentiate amongst each other. The other binary classifications have surprisingly high performances in (a) and (b), noting particularly the success of the  $L(7, 1)$  vs  $L(7, 2)$  classification which are notoriously difficult manifolds to differentiate (they are homotopy equivalent but not homeomorphic). To exemplify the learning behaviour, for this classification the accuracy performance measure was tracked throughout the training process on the training data used for updating the NN parameters and the independent validation data. The accuracy

<sup>16</sup> More specialised architectures may well perform better, but we leave this experimentation to future work, since here we are performing the first tests of these methods for this problem.

<sup>17</sup> The diagonal cases  $M_{ii}$  do not represent sensible experiments so had values set to 1, whilst the symmetric nature of the investigations was enforced by setting the lower diagonal equal to the upper diagonal.



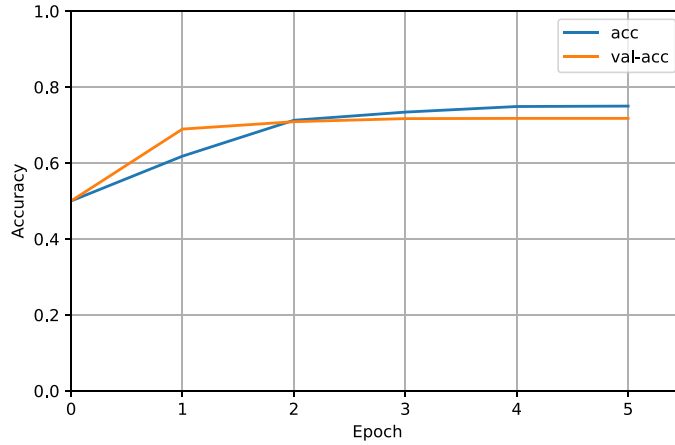
**Figure 11.** Matrices and their corresponding heat maps representing the (a) mean accuracy and (b) mean MCC scores for the binary classification between the data-sets of length 30 IsoSigs generated using  $\{2-3, 3-2\}$  moves; averaged over the 5-fold cross-validation runs. Lighter colours represent higher performance scores, where accuracy evaluates in the range  $[0, 1]$ , and MCC in the range  $[-1, 1]$ . The matrix is manifestly symmetric as the pair order is irrelevant, and the diagonals are trivially 1 where no classification was required. The manifolds are ordered:  $(S^3, S^2 \times S^1, \mathbb{RP}^3, L(7, 1), L(7, 2), T^3, \text{PHS}, H_{\text{SC}})$ .

scores throughout training, evaluating at each epoch, are shown in figure 12. This plot highlights that the performance improves quickly as the general behaviour can be learnt, but does plateau at the limiting score—indicating there are triangulations which are truly difficult to differentiate the manifold structure between.

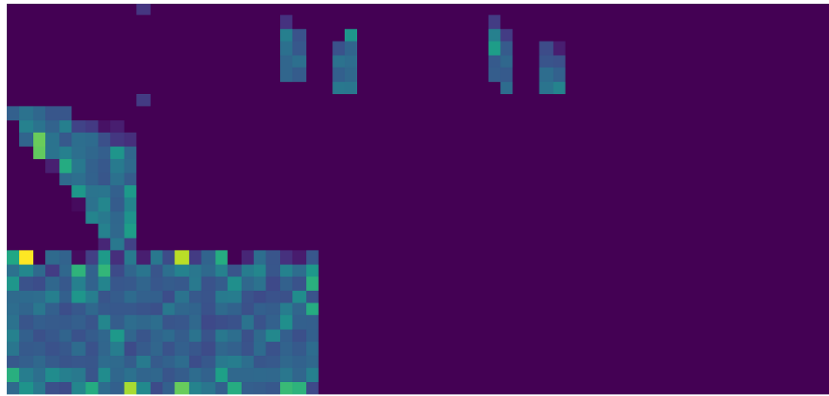
Interestingly, the manifold which is best distinguished from all the other 7 is the Weeks manifold, the only hyperbolic one in the lot. At the opposite,  $S^3, S^2 \times S^1$  and  $\mathbb{RP}^3 = L(2, 1) = SO(3)$  are not distinguished at all.

**4.1.1.1. NN gradient saliency.** We performed an in-depth analysis of the gradient saliency values used by our NNs to distinguish  $L(7, 1)$  vs  $L(7, 2)$ . A trained NN represents a function from inputs to outputs, in the case of these problems this is an input of a single IsoSig, one-hot encoded from length 30 in a length 64 alphabet to a length 1920 vector and an output of a length 2 vector with each entry evaluating in the range  $[0, 1]$  (as a probability distribution over the 2 classes) whose larger value indicates the predicted manifold the input IsoSig triangulation represents.

Computing the derivatives of the NN function outputs with respect to each of the inputs (i.e. for each index of each encoded letter in the IsoSig) provides a gradient function for each input. Evaluating these functions on each IsoSig in the test data gives numerical values for how sensitive the output classification is to each input value. Averaging over all inputs in the test data, and across multiple cross-validation runs, then normalising the output gradients, provides relative gradient measures with greater statistical confidence.



**Figure 12.** Example NN binary classification training curve, between length 30 IsoSigs for manifolds  $L(7, 1)$  and  $L(7, 2)$  generated using  $\{2-3, 3-2\}$  moves up to depth 8. The plot shows that the accuracy performance measure for both training and validation sets increased quickly, converging to the final performance within a few epochs.



**Figure 13.** Image representation of the gradient saliency results for NN binary classification between IsoSigs from manifolds  $L(7, 1)$  and  $L(7, 2)$  generated using  $\{2-3, 3-2\}$  moves. Each row is the encoding of one of the 30 IsoSig characters in the length 64 alphabet, reshaping the 1920 length input vector into a  $(30, 64)$  size matrix. Lighter colours indicate a more significant sensitivity to that input for the output classification.

For the  $L(7, 1)$  vs  $L(7, 2)$  classification, these gradient saliency values were averaged over 100 cross-validation runs, and the final gradient values are represented in figure 13, where each of the 30 rows is an IsoSig character one-hot encoded into the length 64 alphabet. Lighter colours indicate more sensitivity of the classification output to that input. The results in the figure show that the most important parts of the IsoSig input are later characters in the IsoSig and that these are more likely to be lower case letters from the start of the alphabet.

The rough behaviour of the figure 13, respects the IsoSig encoding structure as described in appendix A. The first row is  $\pi(n)$ , and as all these triangulations have 10 tetrahedra only the 10th character in the alphabet (j) has a non-zero saliency and a slightly brighter colour. The

next rows show importance on characters deeper in the alphabet, matching the triple encoding structure of the type sequence where higher index characters are regularly expressed. Following this is a wide-diagonal-like structure up to the 10th character, which coordinates with the destination sequence being bound by  $n$ , which here is 10. Since the tetrahedra are numbered as the destination sequence is built this explains why higher index characters are more important later on as the earlier new tetrahedra seen would be allocated lower numbers. The final section of figure 13 shows a box of important values, up to the 23rd character, matching the permutation sequence behaviour where there're only 23 permutations in  $S_4$  for relabelling vertices. This part shows that all potentially observed permutations are important for the architecture's classification, and since the brightest colour are exhibited in this part this information tends to be the most useful for distinguishing triangulations of these geometries.

Further examining the actual saliency values (represented by colours in figure 13), averaged over all of the cross-validation runs as well as each runs test dataset, we can plot a histogram of the average saliency size, as shown in figure 14(a). This image shows a partition of values with a selection having normalised absolute gradients  $> 10^{-4}$  and the majority having values  $< 10^{-6}$ . There is therefore a clear selection of 'important' inputs and 'non-important' ones. Taking the input indices corresponding to the larger 'important' saliency values  $> 10^{-4}$ , these can be sorted into what character in the alphabet they correspond to (as plotted in a histogram in figure 14(b)), or into what character in the length 30 IsoSig they correspond to (as plotted in a histogram in figure 14(c)).

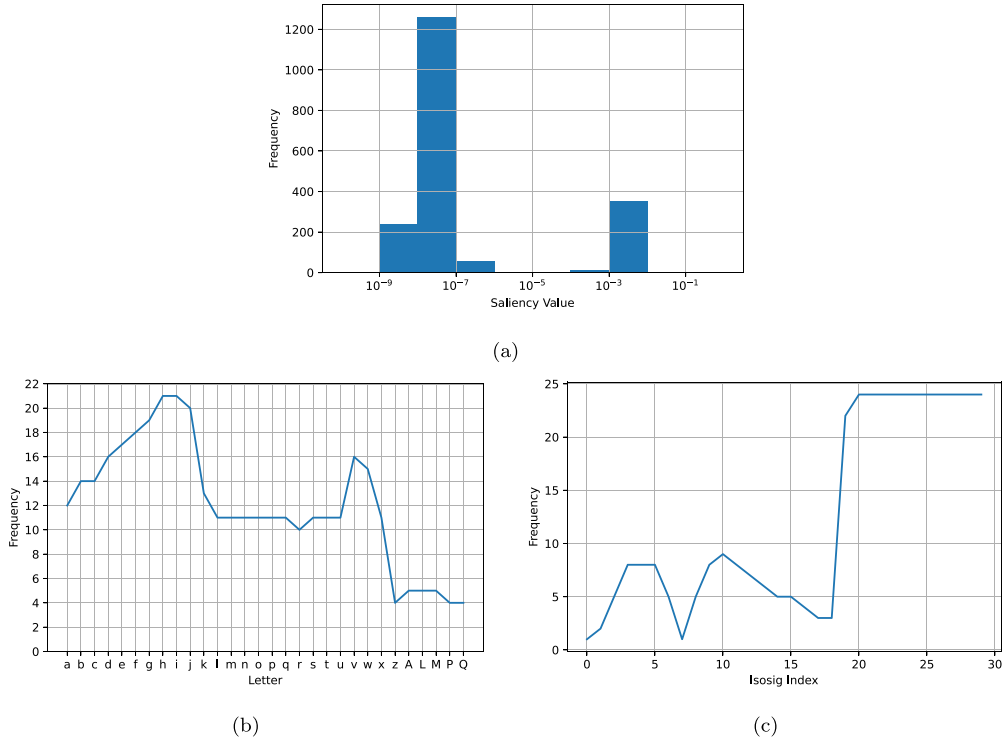
Figure 14(b) shows that the 363 'important' inputs correspond to only a subset of the full alphabet—predominantly the beginning of this alphabet with a few letters deeper in. Equivalently, these same important input features have much higher frequencies towards the end of the IsoSig as shown in figure 14(c). From this saliency analysis the NNs are hence relying most on the final entries of the IsoSigs (the permutation sequence) to determine the manifold, and the value of these entries becomes discriminant. Additionally the occurrence of specific letters (like *h* or *i* with the highest frequency in figure 14(b)) can significantly help the architectures' confidence in classification.

**4.1.2. Transformers.** Since the IsoSig data closely resembles that of text, one may sensibly question how transformer architectures, which have seen great performance in many natural language processing problems [79, 80], perform when classifying this data.

A popular baseline for performing binary classification tasks with transformers is to use the BERT (Bidirectional Encoder Representations from Transformers) architecture [81]. In a similar vein to that employed in section 4.1.1, the BERT model is now used to classify whether a given pair of IsoSigs represent the triangulations of the same manifold. This model is implemented using the `transformers` library [82].

Transformers are primarily successful as large language models, and due to this tend to have a very large number of trainable parameters. To make training feasible, their use often relies upon the principle of *transfer learning*, where the models are pre-trained on a large standard set of data, and then retrained on a smaller specialised set. This use of pre-trained models can avoid tens of thousands of core hours in retraining, and is what makes these large language models feasible to use at all. However, the pre-trained models publicly available are trained on language data (often English language text). This hence, raises appropriate concerns about the success of transfer learning onto this IsoSig 3-manifold 'language' data, which does not resemble the syntactical and grammatical structure of the English language.

Nevertheless, due to their paramount successes in recent times, it is interesting to examine their performance in this context. To do so we take the binary classification of the  $L(7, 1)$



**Figure 14.** Saliency analysis for the  $L(7, 1)$  vs  $L(7, 2)$  binary classification, averaged over 100 NNs. A histogram of the average saliency absolute values across the 1920 input vectors is shown in (a), demonstrating a partition into more/less important features. Subsequently (b) shows a histogram of the more important features with saliency value  $> 10^{-4}$  sorted according to the letter in the alphabet they correspond to; whilst (c) shows a histogram of the same set sorted according to which of the 30 characters in the input IsoSig they correspond to.

and  $L(7, 2)$  datasets, as performed by the NNs to Accuracy and MCC scores of (0.705, 0.409) respectively, and repeat the training and testing using the BERT model. Using a single layer output classifier to change the standard 768 output dimension of the BERT model to the 2 binary classes, along with equivalent hyperparameters such as training with the Adam optimiser and cross-entropy loss for 30 epochs with a batch size of 64 and learning rate of 0.001. The transformer performance scores were:

$$\text{Accuracy} = 0.689, \quad (4.4)$$

$$\text{MCC} = 0.381. \quad (4.5)$$

Unfortunately, whilst some learning was achieved, the performance was not as good as the NN models, likely due to the inappropriate nature of the English language pre-training misleading the classification. In addition to worse performance, the transformers also required significantly more computational resources (time, power, memory). To exemplify this, whilst the NNs took the order of seconds to train and test, this simplest transformer model took the order of hours, and required orders of magnitude more memory to run.

**Table 4.** The knot IsoSigs used to start the generation of Pachner graphs using only  $2-3$  and  $3-2$  moves. Each knot was used to construct a knot-complement 3-manifold, which was then triangulated to give a 1-vertex triangulation. The respective IsoSigs are listed in this table, along with the number of tetrahedra in this initial triangulation.

| Knot                 | Initial IsoSig   | # Tetrahedra |
|----------------------|------------------|--------------|
| Unknot               | cMcabbgds        | 2            |
| Trefoil              | cPcbbbadu        | 2            |
| Figure eight         | cPcbbbiht        | 2            |
| $5_1$                | dLQbcccaekv      | 3            |
| $5_2$                | dLQbcccdero      | 3            |
| DT $6a_3 = 6_1$ knot | eLPkbcdddcwjb    | 4            |
| DT $8n_1 = 8_2$ knot | fLLQcbcdeedowxxd | 5            |

Therefore, although it was relevant experimentation, these results support the motivation to focus on the NN architecture for learning of triangulation IsoSig properties.

#### 4.2. Differentiating Knots

By a famous result of Gordon and Luecke, the complement of a knot entirely determines the knot [83]. This allows one to identify knots with 3-manifolds, as the knot complement embedded within  $S^3$ ; and with this motivation the workflow we have set-up may also be directly extended to study and distinguish knots.

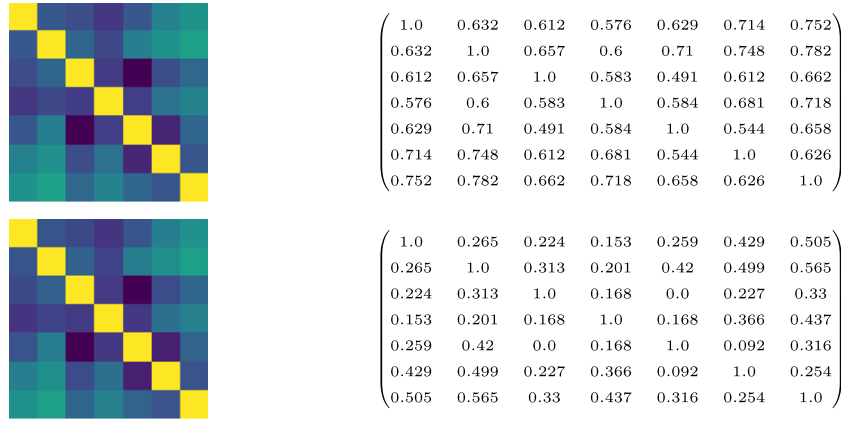
In this work, we again put focus on a selection of knots, in particular focusing on the most fundamental knots with a low number of crossings. These knots are labelled: {Unknot, Trefoil, figureEight,  $5_1$ ,  $5_2$ , DT  $6a_3$  (a.k.a.  $6_1$ ), DT  $8n_1$  (a.k.a.  $8_2$ )}, where the final knot is hyperbolic non-alternating. The figure eight and the  $5_2$  knot are the smallest hyperbolic knots in the 3-sphere and indeed it turns out that they are more difficult to distinguish. In a similar way one could expect some difficulties in distinguishing the torus knots  $3_1$  (the trefoil) and  $5_1$  (the cinquefoil) but it is not so much the case (see figure 15).

Again all the initial IsoSigs are 1-vertex, and are generated using only  $2-3$  and  $3-2$  moves. The initial IsoSigs for these knot complement 3-manifolds are given in table 4, along with the respective number of tetrahedra in each initial triangulation.

The Pachner graphs were generated up to a depth that provided a suitable amount of data for a fixed IsoSig length. Again searching for 2000 IsoSigs per Pachner graph, the selected IsoSig length was 25 characters. From there the same NN architecture was trained to perform binary classification between the knot complement manifolds. The binary classification learning was performed for each pairing of knot-complement 3-manifolds, where the accuracy and MCC scores are displayed in figure 15.

The results show comparable performances between knots, however with worse maximum performances compared to the general 3-manifold learning in section 4.1. Some pairings had close to no learning, better exemplified by the near-zero MCC scores for the (figureEight,  $5_2$ ) and the ( $5_2$ , DT  $6a_3$ ) pairings. Generally, as shown by the images, the colours are darker closer to the diagonal, indicating worse performance, and suggesting that in general the more disparate the crossing structure of the knot the easier the NNs can distinguish them. Accuracies for identifying the Unknot are not as high as in the focused Unknot study in [44].





**Figure 15.** Matrices and corresponding heat maps representing the (a) mean accuracy and (b) mean MCC scores for the binary classification between the datasets of length 25 knot complement IsoSigs generated using  $\{2-3, 3-2\}$  moves; averaged over the 5-fold cross-validation runs. Lighter colours represent higher performance scores, where accuracy evaluates in the range  $[0, 1]$ , and MCC in the range  $[-1, 1]$ . The matrix is manifestly symmetric as the pair order is irrelevant, and the diagonals are trivially 1 where no classification was required. The manifolds are ordered: (Unknot, Trefoil, figure eight,  $5_1$ ,  $5_2$ ,  $DT\ 6a_3$ ,  $DT\ 8n_1$ ).

**4.2.1. Supplementing with Dehn Surgery.** In addition to learning with their knot-complement manifolds, we also perform two Dehn surgery operations (0-surgery and 1-surgery) whereby a tubular neighbourhood of a knot is drilled out and then filled with a solid torus [84].

Dehn surgery is a standard operation in topology allowing modification of the topology of a 3-manifold by a ‘cut and paste’ operation on the  $D^2 \times S^1$ -neighborhood of a knot consisting in re-gluing the solid torus as  $S^1 \times D^2$ . This operation is also called 0-Dehn surgery and more in general  $n$ -Dehn surgery consists in regluing the solid torus after twisting the image of the meridian  $n$ -times along the longitude. In particular operating 0 or 1 surgeries along knots in  $S^3$  one produces closed oriented 3-manifolds with the homology respectively of  $S^2 \times S^1$  and of  $S^3$ . These manifolds are themselves canonically associated to the knots so their homeomorphism type is by itself a knot invariant, albeit a complicated one. Here we try and distinguish knots by distinguishing these surgery manifolds.

Focus is put on the differentiation between the Unknot and Trefoil binary classification problem, which for the knot complement 3-manifolds classified with accuracy 0.632 and MCC 0.265, neither particularly impressive scores. In this extension, the initial knot complement triangulation IsoSigs, as given in table 4, had either 0-surgery or 1-surgery performed on them, to produce further 3-manifolds whose IsoSigs are given in table 5. We remind the reader that the 0 and 1 surgeries on the Unknot are respectively  $S^2 \times S^1$  and  $S^3$  and on the left-handed Trefoil they are two Seifert fibered manifolds, respectively  $(S^2, (2, 1), (3, 1), (6, -5))$  and  $(S^2, (2, 1), (3, 1), (7, -6))$  (the latter *not* being the Poincaré homology sphere, which is the 1 surgery over the *right*-handed Trefoil). Each of these was used to seed Pachner graph generation to generate databases of 2000 length 30 IsoSigs. Noting that in this subsection length 30 was used again instead of the length 25 as used in section 4.2, this is because the initial

**Table 5.** The surgery Isosigs used to start generation of Pachner graphs using only  $2 - 3$  and  $3 - 2$  moves.

| Manifold | Surgery | IsoSig                 | # Tetrahedra |
|----------|---------|------------------------|--------------|
| Unknot   | 0       | cMcabbja j             | 2            |
|          | 1       | cMcabbgag              | 2            |
| Trefoil  | 0       | gvLQQcdefeffnjndsp x   | 6            |
|          | 1       | hLLLQkcdefgfgghsdaenjw | 7            |

**Table 6.** NN binary classification performance scores for the knot complements, the 0-surgeries, and the 1-surgeries, between the Unknot and Trefoil knots. Scores were averaged over 5 cross-validation runs.

| Investigation   | Accuracy | MCC   |
|-----------------|----------|-------|
| Knot complement | 0.632    | 0.265 |
| 0-surgery       | 0.935    | 0.871 |
| 1-surgery       | 0.974    | 0.949 |

IsoSigs after surgery lead to length distributions with very few length 25 IsoSigs<sup>18</sup>. Another comment worth emphasising is that whilst the IsoSigs in table 5 have a range of initial number of tetrahedra, since only a fixed length IsoSig is taken from the graph for ML these IsoSigs will also correspond to triangulations with the same number of tetrahedra. Additionally, it is worth noting that the initial IsoSig from 0-surgery of the Unknot is a triangulation with 3-vertices (whilst the others are still all 1-vertex), this means this Pachner graph is no longer a 1-vertex Pachner graph.

With these datasets of 2000 length 30 IsoSigs corresponding to surgeries of these knot complements, the respective ML binary classification between the knots was repeated for each surgery operation. The NN architecture used was the same as throughout this work, as quoted in section 4.1.1. The 5-fold cross-validation performance scores are given in table 6, including the values for the knot complements without surgery repeated from section 4.2.

Performances are substantially higher after surgery, for either surgery actions. It is interesting to observe that since 0 and 1-surgeries over the unknot are  $S^2 \times S^1$  and  $S^3$  respectively, the problem discussed here is yet another instance of the problem analysed in section 4.1 and the accuracies for the 0-surgery and 1-surgery cases are comparable with those seen in figure 11. These performances are quite high and show that one can easily complement the initial performances for the simple knot complements with further parallel analyses. The potential for distinguishing knots seems therefore quite high.

## 5. Conclusions and outlook

In this work triangulations of 3-manifolds were studied via their IsoSig representation. A selection of 3-manifolds were chosen to be of focus in this study, along with a selection of knot complements from knots exhibiting a low number of crossings. Their databases of triangulations were generated through their Pachner graphs, exhaustively performing  $\{2 - 3, 3 - 2\}$

<sup>18</sup> This is also why just these two simplest knots are considered since the length distributions for some knot complements after surgery did not even include any length 25 or 30 IsoSigs.

moves on an initial triangulation up to some pre-specified depth. The graph structure of these graph was also analysed through a variety of network analysis techniques.

Pachner graph network analysis made some preliminary connections between the 3-manifold topological properties and how their respective Pachner graph grows, particularly using eigenvector centrality and decomposition of their minimum cycle bases. This included proposal of a conjectural relation (conjecture 1) between the growth rate of the Pachner graph and the systole of a hyperbolic 3-manifold. ML results showed that through a one-hot encoding of the IsoSigs that simple NN architectures could learn to well distinguish which manifold a triangulations' IsoSig corresponded to—a feat very difficult by eye due to the high information density of this encoding scheme. Gradient saliency analysis then demonstrated that the NNs were using the IsoSig characters in a manner which respected the encoding schemes decomposition into type, destination, and permutation sequences.

Transformers could not reach the same performance scores as the NNs, likely due to the highly unique language-style of this representation. However, the NN architectures could generalise well to differentiating these IsoSig representations of knot complements, and particularly well manifolds after the operation of Dehn surgery.

At this work's respective [GitHub](#), code functionality is provided for generating and visually representing Pachner graphs for generic 3-manifolds; as well as performing a range of network analysis techniques. Scripts for repeating the ML investigations performed here are also provided, as well as databases of IsoSigs for the use of interested readers.

Future work would aim to make the connections between Pachner graph network properties and manifold topological properties more concrete. Further to this, ML methods may be applied directly to the Pachner graphs, approximating optimal search techniques for problems such as minimising number of tetrahedra in a representation, or conversely using graph-neural-network techniques to study them. Additionally the study of more manifolds, and of triangulations deeper into the Pachner graphs would be interesting to corroborate results observed here.

## Data availability statement

The data that support the findings of this study are openly available at the following URL/DOI: <https://github.com/edhirst/IsoSigPGML>.

## Acknowledgment

F C was supported by CIMI Labex ANR 11-LABX-0040 at IMT Toulouse within the program ANR-11-IDEX-0002-02. YHH would like to thank STFC for Grant ST/J00037X/2 and a Leverhulme project grant. E Heyes would like to thank SMCSE at City, University of London for the PhD studentship, as well as the Jersey Government for a postgraduate grant. E Hirst acknowledges support from Pierre Andurand over the course of this research. This research utilised Queen Mary's Apocrita HPC facility [68], supported by QMUL Research-IT.

## Appendix A. IsoSig encoding

Isomorphism Signatures, or IsoSigs, as introduced in [85], are a way of representing 3-manifold triangulations. In addition to uniquely representing each isomorphism class, they are both low memory and fast to compute/check/search.

The algorithm to produce an IsoSig for a given 3-manifold triangulation with  $N$  tetrahedra starts with the generation of *canonical labellings*. Before defining a canonical labelling we will first define a labelling, which is generated as follows:

1. Arbitrarily label the tetrahedra from 0 to  $N - 1$ .
2. Arbitrarily label each tetrahedra's four vertices from 0 to 3. Respectively the face opposite each vertex  $i$  is given the same label  $i$ .
3. Using the notation  $A_{t,f}$  to equal the label of the tetrahedra glued to face  $f$  on tetrahedra  $t$  (such that  $A_{t,f} \in \{0, \dots, N - 1\}$ ); list the gluings as:  
( $A_{0,0}, A_{0,1}, A_{0,2}, A_{0,3}, A_{1,0}, \dots, A_{N-1,3}$ ) to define the labelling.

There are many possible labellings that uniquely define a triangulation, we first reduce this set to only consider canonical labellings, which are defined as labellings with the properties:

- Tetrahedron  $t$  first appears in the labelling before tetrahedron  $t'$  iff  $t < t'$ , for  $t, t' \geq 1$  since 0 is implicitly first.
- Where tetrahedron  $t$  first appears in the list at position  $A_{t',f'}$ , the labelling of tetrahedra  $t$ 's vertices matches the labelling of tetrahedra  $(t')$ 's—i.e. each of the three vertices in face  $f'$  of tetrahedra  $t$  has the same vertex labelling as the vertex it is glued too in tetrahedra  $t'$ , the remaining fourth vertex labelling is then inferred from these three.

For any triangulation, a canonical labelling can be directly generated by first selecting any of the  $N$  tetrahedra as the 0th tetrahedron; then arbitrarily labelling its vertices (of which there're  $4! = 24$  ways). From this choice the remainder of the canonical labelling can be inferred by following the gluings list—each time a new tetrahedron is introduced label it as the next one in the list and label its vertices using the identity map such that its vertex labels match the current tetrahedron in accordance with the canonical labelling definition above. Through this labelling process, it can be seen that there're  $24N$  canonical labellings for any triangulation [85].

This process of allocating a canonical labelling is clearly still redundant. To resolve this many-to-one map from labellings to triangulations, all canonical labellings are computed, encoded as IsoSigs, then the lexicographically smallest encoding is selected to represent the triangulation. How the labellings are encoded is what gives the IsoSigs their character-based / language-like representation, as well as being significantly more efficient.

A first step in improving the encoding efficiency is using an alphabet, and when the number of tetrahedra  $N$  is large this becomes particularly effective. The alphabet<sup>19</sup> implemented is in base 64, such that:

|           |   |     |    |    |     |    |    |     |    |    |    |
|-----------|---|-----|----|----|-----|----|----|-----|----|----|----|
| Integer   | 0 | ... | 25 | 26 | ... | 51 | 52 | ... | 61 | 62 | 63 |
| Character | a | ... | z  | A  | ... | Z  | 0  | ... | 9  | +  | -  |

Redundant information in the gluings list can also be omitted. In this process, the gluings list through the triangulation faces  $A_{t,f}$  is re-encoded into three parts: (i) destination sequence, (ii) type sequence, (iii) permutation sequence. Next we describe each of these part individually:

<sup>19</sup> To encode larger numbers  $d = \lfloor \log_{64}(N) \rfloor + 1$  characters can be used by just converting all numbers to base 64, fixing  $d$  characters for all encoded integers, and using padding when necessary. To exemplify, for  $N = 80$ , then  $d = 2$  characters are needed, and the number 12 would be written as (12, 0) in the  $(64^0, 64^1)$  basis, such that the encoding would be `ma`.

- (i) Destination Sequence: Since each face gluing involves two faces, the gluings list for each face naturally includes the information about each gluing twice. Therefore, from the gluings list  $(A_{0,0}, A_{1,0}, \dots, A_{N-1,3})$ , each time a face  $A_{t,f}$  is glued to another face  $A_{t',f'}$  which has already been listed (i.e.  $t' < t$ , or  $t' = t$  and  $f' < f$ ), the entry  $A_{t,f}$  is omitted from the gluings list. In addition for more general triangulations with boundary, a character  $\partial$  may be used instead of the tetrahedron number for  $A_{t,f}$  to evaluate to, when the face  $f$  is a boundary of the triangulation. The destination sequence is then a list of all the unique gluings (and boundaries).
- (ii) Type Sequence: Each term in the destination sequence is then classified as one of three types. Either type 0  $\rightarrow$  the face is a boundary ( $A_{t,f} = \partial$ ); type 1  $\rightarrow$  the face is glued to a new tetrahedron; or type 2  $\rightarrow$  the face is glued to a previously seen tetrahedron. The type sequence is then a list of the numbers 0, 1, or 2 for each term in the destination sequence.
- (iii) Permutation Sequence: For each non-boundary face, the gluing information is described by a permutation of the vertices from the face in the first tetrahedron to the face in the second (each of the three vertices in the face are appropriately paired, and the remaining fourth vertex in each case can then be inferred). Indexing all the 24 possible permutations  $p_{t,f} \in S_4$  from 0 up to 23 (i.e.  $(0, 1, 2, 3) := 0$ ,  $(0, 1, 3, 2) := 1$ , etc), the permutation sequence is then a number for each non-boundary term in the destination sequence.

The full IsoSig is then a further reduction of this information. The map  $\pi(\cdot)$  takes any integer into its alphabet encoding; if  $N < 63$  the IsoSig starts with  $\pi(N)$ ; when  $N \geq 63$ , the sequence starts<sup>20</sup> with  $\pi(63)\pi(d)\varepsilon(N)$ , where  $\varepsilon(\cdot)$  encodes each of the digits of  $N$  written modulo 64 (so when  $N < 64$  the  $d = 1$  encoding is implicit).

Next the type sequence is encoded and listed, collecting three consecutive terms into each printable character via  $\pi(\tau_i + 4\tau_{i+1} + 16\tau_{i+2})$  (since each integer  $\tau \in \{0, 1, 2\}$ ), padding the end with zeros if necessary. Then the destination sequence is encoded and listed via  $\pi(A_{t,f})$ , but *only* for the type 2 faces (since type 0 are boundary, and type 1 can be deduced as the next tetrahedra label since the labelling is canonical). Finally, the permutation sequence is encoded and listed, but again *only* for the type 2 faces (since the type 0 faces have no gluing, and the type 1 faces are implicitly the identity permutations since the labelling is canonical).

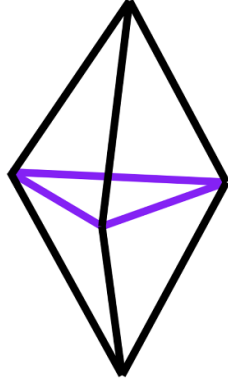
This produces the encoding of the canonical labelling, however as previously mentioned there are  $24N$  of these. To uniquely specify the IsoSig, each of these encoded canonical labellings is computed, and the lexicographically smallest is selected as the IsoSig. The lexicographical ordering used is the ASCII computational standard [86].

### A.1. Example: triangular bipyramid

As a simple example let us consider the triangular bipyramid, as depicted in figure 16. This has two tetrahedra, which is far less than 63, so the encoding begins with  $c = \pi(2)$  (the  $d = 1$  is implicit).

Arbitrarily starting with the top tetrahedron as our tetrahedron 0, let us choose the vertex labelling such that the single glued face has vertices 0, 1, and 2; with the remaining vertex labelled 3 (which matches the labelling of the glued face as it is opposite). The gluings list

<sup>20</sup> Since 63 is the highest number which can be encoded if the sequence starts  $\pi(63)$  it is considered as ‘63 or above’, then the next character is taken to be  $\pi(d)$  followed by the full  $\pi(N)$  (which may be 63 again).



**Figure 16.** The triangular bipyramid, an example 3-manifold triangulation, where the coloured edges indicate the glued face.

then reads  $A_{0,0} = \partial$ ,  $A_{0,1} = \partial$ ,  $A_{0,2} = \partial$ ,  $A_{0,3} = 1$ , now this is the first occurrence of this new tetrahedron (bottom) and hence sets its tetrahedron label to the next integer which is 1; furthermore the canonical nature of the labelling sets the vertex labellings for this second tetrahedron, each vertex on the glued face in tetrahedron 0 is glued to a vertex of the same labelling in tetrahedron 1. The gluings list hence continues  $A_{1,0} = \partial$ ,  $A_{1,1} = \partial$ ,  $A_{1,2} = \partial$ ,  $A_{1,3} = 0$ . The overall gluings list is then:  $(\partial, \partial, \partial, 1, \partial, \partial, \partial, 0)$ .

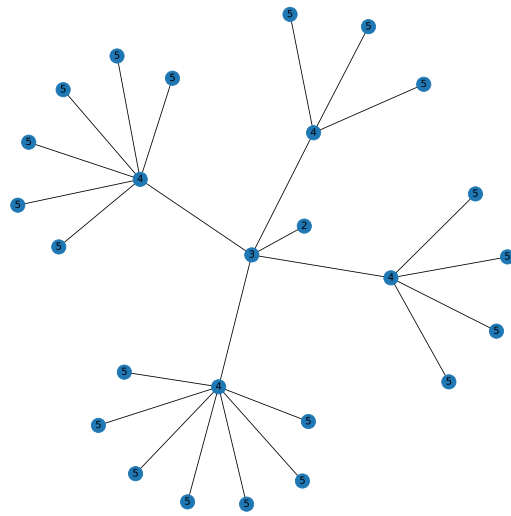
This reduces to the destination sequence:  $(\partial, \partial, \partial, 1, \partial, \partial, \partial)$ , removing the repetition of the initial tetrahedron at the end of the gluings list. The type sequence is then defined as:  $(0, 0, 0, 1, 0, 0, 0)$  over the destination sequence, i.e. there are no non-trivial gluings (type 2's). Finally the permutation sequence is computed, which is only defined for non-boundary terms in the destination sequence:  $(0)$ .

These sequences are then encoded and combined, after the initial  $c$ , the type sequence is then split into threes and padded as  $((0, 0, 0), (1, 0, 0), (0, 0, 0))$ , becoming  $\pi(0 + 4 \times 0 + 16 \times 0) = \pi(0) = a$ ,  $\pi(1 + 4 \times 0 + 16 \times 0) = \pi(1) = b$ ,  $\pi(0 + 4 \times 0 + 16 \times 0) = \pi(0) = a$ , which together is  $aba$ . Following this, since there are no type 2 faces, the destination and permutation sequences are trivially empty, and this defines the full encoded canonical labelling as  $caba$ , which to check is canonical we would have to consider all  $24 \times 2 = 48$  canonical labellings and choose the lexicographically smallest (which in this case it is!).

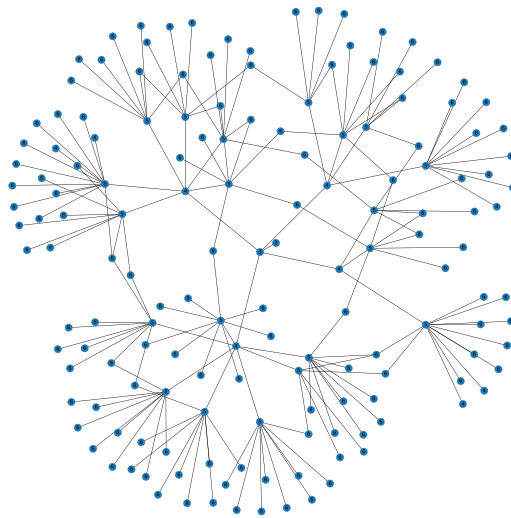
## Appendix B. Pachner graphs

The Pachner graphs generated with moves  $\{2-3, 3-2\}$  up to depths 3 and 4, for the remaining 8 considered manifolds, starting from triangulations with IsoSigs specified in table 1. The nodes represent triangulations and are labelled with the triangulation's respective number of tetrahedra, edges represent Pachner moves which transform the triangulations they connect into each other.

The smallest closed hyperbolic manifold  $H_{SC}$  had too many IsoSigs in its Pachner graph to plot at even depth 3, therefore the depths 1 and 2 are shown instead. To exemplify the speed of Pachner graph growth, the depth 5  $S^3$  Pachner graph is also given at the end. These are given in figures 17–31.

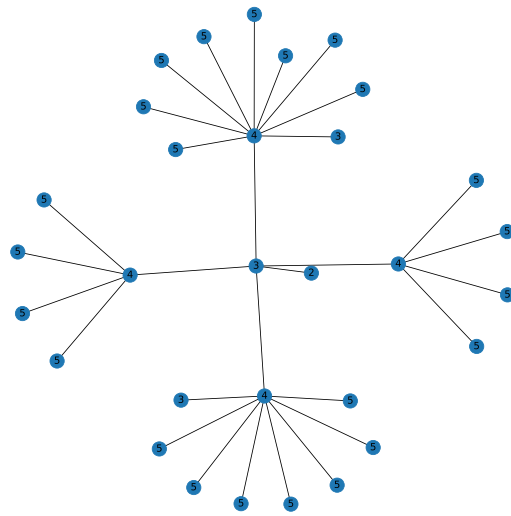


**Figure 17.**  $S^2 \times S^1$  Depth 3.

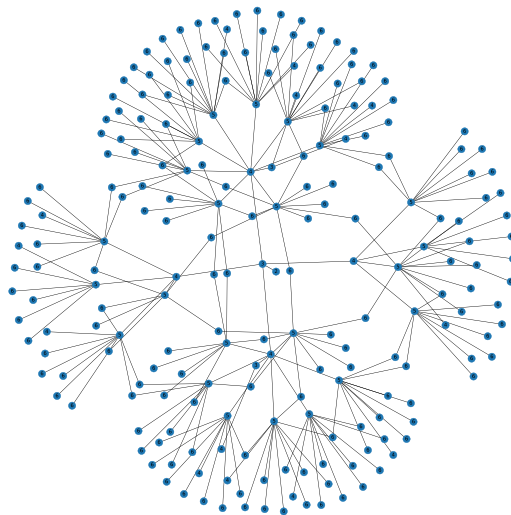


**Figure 18.**  $S^2 \times S^1$  Depth 4.

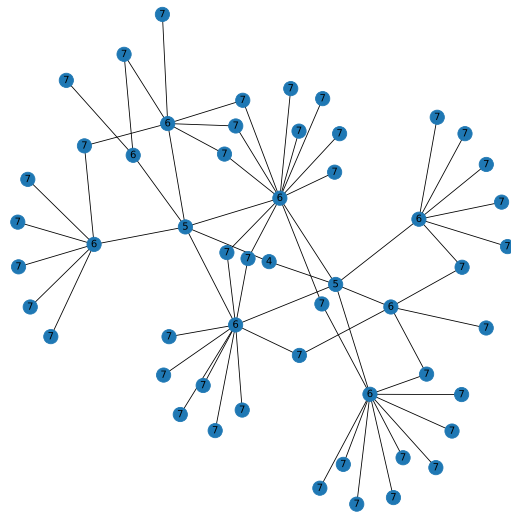




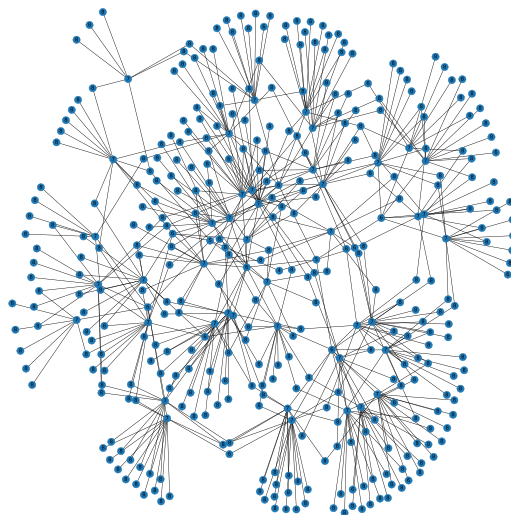
**Figure 19.**  $\mathbb{RP}^3$  Depth 3.



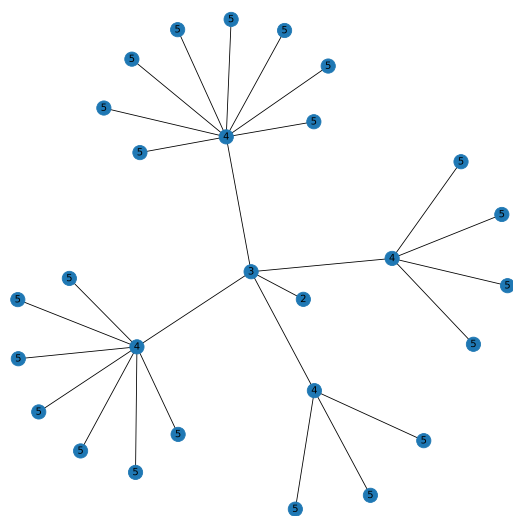
**Figure 20.**  $\mathbb{RP}^3$  Depth 4.



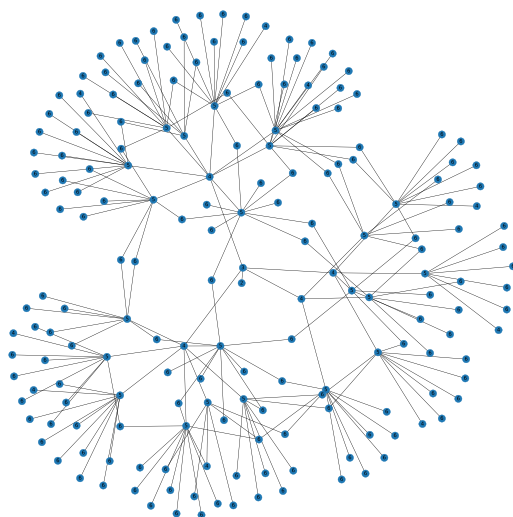
**Figure 21.**  $L(7, 1)$  Depth 3.



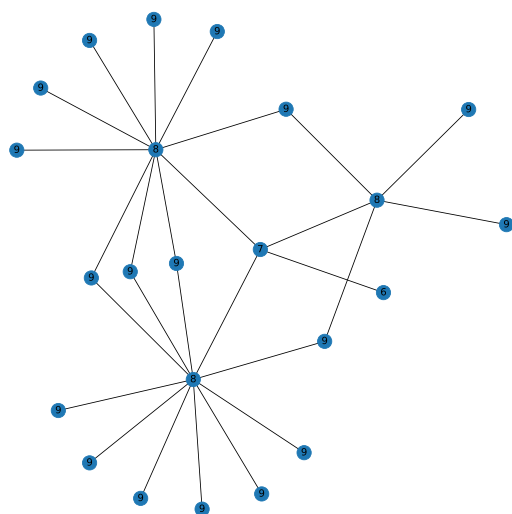
**Figure 22.**  $L(7, 1)$  Depth 4.



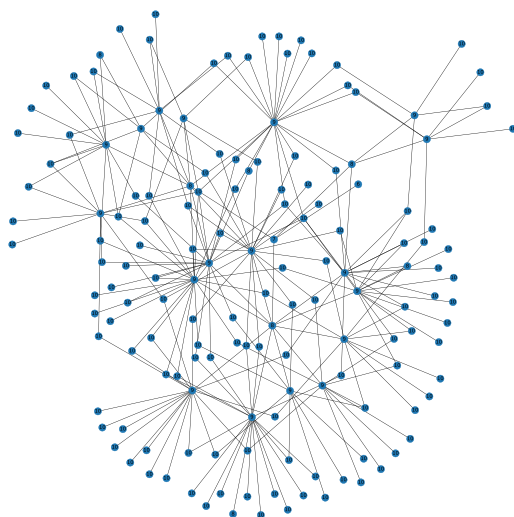
**Figure 23.**  $L(7,2)$  Depth 3.



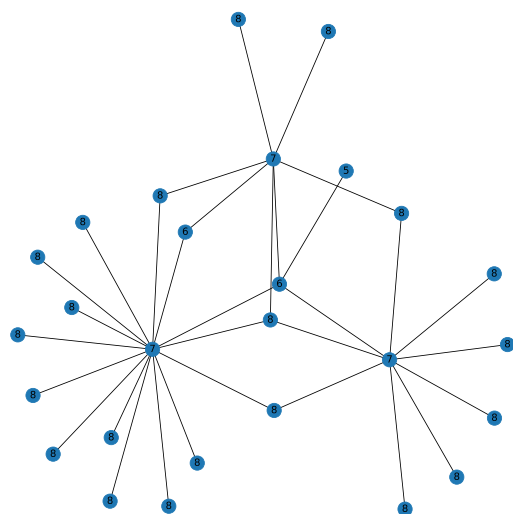
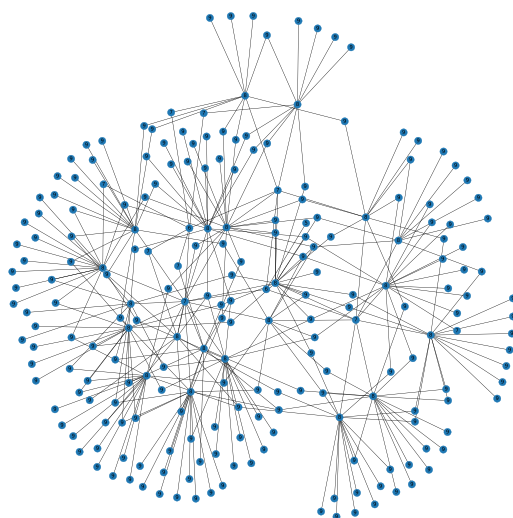
**Figure 24.**  $L(7,2)$  Depth 4.

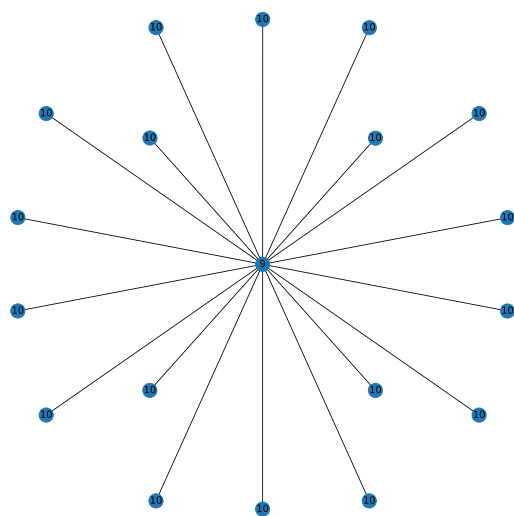


**Figure 25.**  $T^3$  Depth 3.

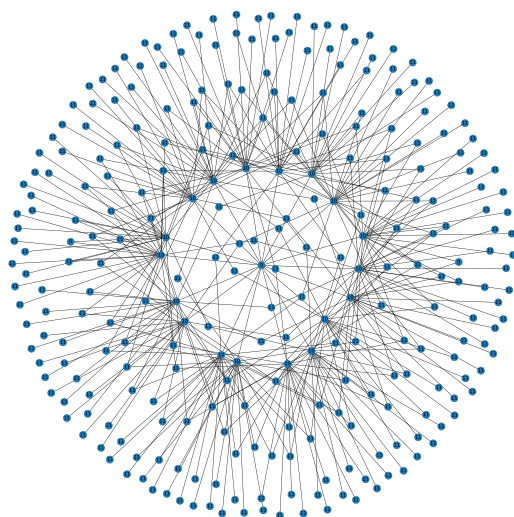


**Figure 26.**  $T^3$  Depth 4.

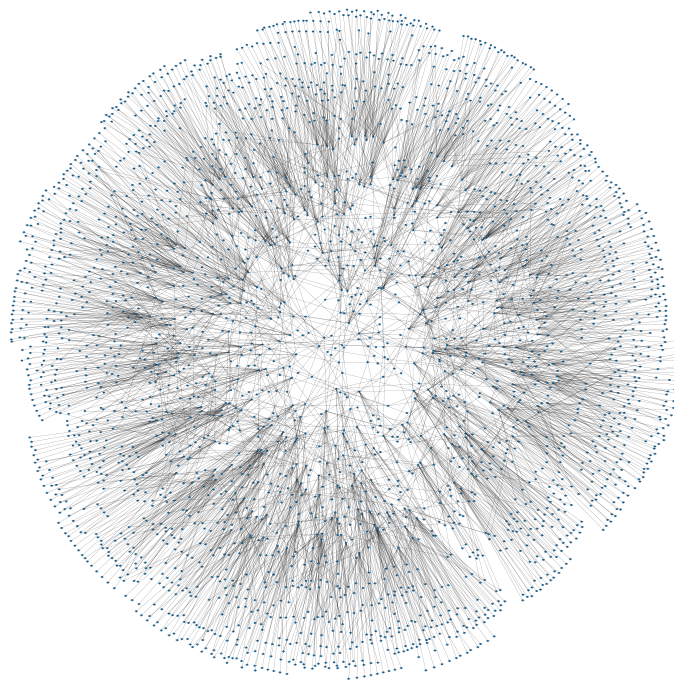
**Figure 27.** PHS Depth 3.**Figure 28.** PHS Depth 4.



**Figure 29.**  $H_{SC}$  Depth 1.



**Figure 30.**  $H_{SC}$  Depth 2.



**Figure 31.**  $S^3$  depth 5.

## ORCID iDs

Yang-Hui He  <https://orcid.org/0000-0002-0787-8380>

Elli Heyes  <https://orcid.org/0000-0002-7608-3485>

Edward Hirst  <https://orcid.org/0000-0003-1699-4399>

## References

- [1] Thurston W P 1982 Three-dimensional manifolds, Kleinian groups and hyperbolic geometry *Bull. Am. Math. Soc.* **6** 357–81
- [2] Perelman G 2006 The Entropy formula for the Ricci flow and its geometric applications (arXiv:[math/0211159](https://arxiv.org/abs/math/0211159))
- [3] Perelman G 2006 Ricci flow with surgery on three-manifolds (arXiv:[math/0303109](https://arxiv.org/abs/math/0303109))
- [4] Witten E 1989 Quantum field theory and the jones polynomial *Commun. Math. Phys.* **121** 351–99
- [5] C N Yang, M-L Ge and Y-H He ed 2019 *Topology and Physics* (World Scientific)
- [6] Zomorodian A J 2005 *Topology for Computing (Cambridge Monographs on Applied and Computational Mathematics)* (Cambridge University Press)
- [7] Brodie C R, Constantin A, Deen R and Lukas A 2020 Machine learning line bundle cohomology *Fortsch. Phys.* **68** 1900087
- [8] Douglas M R, Lakshminarasimhan S and Qi Y 2020 Numerical Calabi-Yau metrics from holomorphic networks (arXiv:[2012.04797](https://arxiv.org/abs/2012.04797) [hep-th])
- [9] Krippendorf S and Spannowsky M 2022 A duality connecting neural network and cosmological dynamics *Mach. Learn. Sci. Tech.* **3** 035011
- [10] Aslan B, Platt D and Sheard D 2023 Group invariant machine learning by fundamental domain projections *NeurIPS Workshop on Symmetry and Geometry in Neural Representations* (PMLR) pp 181–218



- [11] Berman D S, Klinger M S and Stapleton A G 2023 Bayesian renormalization *Mach. Learn. Sci. Tech.* **4** 045011
- [12] Halverson J and Ruehle F 2023 Metric flows with neural networks (arXiv:2310.19870 [hep-th])
- [13] Alawadhi R, Angella D, Leonardo A and Gherardini T S 2024 Constructing and machine learning Calabi-Yau five-folds *Fortsch. Phys.* **72** 2300262
- [14] Butbaia G *et al* 2024 Physical yukawa couplings in heterotic string compactifications (arXiv:2401.15078 [hep-th])
- [15] He Y-H, Hirst E and Peterken T 2021 Machine-learning dessins d'enfants: explorations via modular and Seiberg–Witten curves *J. Phys. A: Math. Theor.* **54** 075401
- [16] Bao J, Franco S, He Y-H, Hirst E, Musiker G and Xiao Y 2020 Quiver mutations, seiberg duality and machine learning *Phys. Rev. D* **102** 086013
- [17] Bao J, He Y-H, Hirst E, Hofschneider J, Kasprzyk A and Majumder S 2022 Hilbert series, machine learning and applications to physics *Phys. Lett. B* **827** 136966
- [18] Bao J, He Y-H and Hirst E 2022 Neurons on Amoebae *J. Symb. Comput.* **116** 1–38
- [19] Bao J, He Y-H, Hirst E, Hofschneider J, Kasprzyk A and Majumder S 2023 Polytopes and machine learning *Math. Sci.* **01** 181–211
- [20] Berman D S, He Y-H and Hirst E 2022 Machine learning Calabi-Yau hypersurfaces *Phys. Rev. D* **105** 066002
- [21] Arias-Tamargo G, He Y-H, Heyes E, Hirst E and Rodriguez-Gomez D 2022 Brain webs for brane webs *Phys. Lett. B* **833** 137376
- [22] Dechant P-P, He Y-H, Heyes E and Hirst E 2023 Cluster algebras: network science and machine learning *J. Comput. Algebra* **8** 100008
- [23] Aggarwal D, He Y-H, Heyes E, Hirst E, Earp H N S and Silva T S R 2024 Machine learning Sasakian and G2 topology on contact Calabi-Yau 7-manifolds *Phys. Lett. B* **850** 138517
- [24] Choi E and Seong R-K 2024 Machine learning regularization for the minimum volume formula of toric Calabi-Yau 3-folds *Phys. Rev. D* **109** 046015
- [25] He Y H, Jejjala V, Mishra C and Sharnoff M 2023 Learning to be simple (arXiv:2312.05299 [cs.LG])
- [26] Hirst E and Gherardini T S 2024 Calabi-Yau four-, five-, sixfolds as  $\mathbb{P}_w^n$  hypersurfaces: machine learning, approximation and generation *Phys. Rev. D* **109** 106006
- [27] Nasteski V 2017 An overview of the supervised machine learning methods *Horizons* **4** 56
- [28] Seong R-K 2023 Unsupervised machine learning techniques for exploring tropical coamoeba, brane tilings and Seiberg duality *Phys. Rev. D* **108** 106009
- [29] Chen S, Dechant P-P, He Y-H, Heyes E, Hirst E and Riabchenko D 2024 Machine Learning Clifford Invariants of ADE Coxeter Elements *Adv. Appl. Clifford Algebras* **34** 20
- [30] Ghahramani Z 2003 Unsupervised learning *Summer School on Machine Learning* (Springer) pp 72–112
- [31] Halverson J, Nelson B and Ruehle F 2019 Branes with brains: exploring string vacua with deep reinforcement learning *J. High Energy Phys.* **JHEP06(2019)003**
- [32] Abel S, Constantin A, Harvey T R and Lukas A 2022 Evolving heterotic gauge backgrounds: genetic algorithms versus reinforcement learning *Fortsch. Phys.* **70** 2200034
- [33] Cole A, Krippendorf S, Schachner A and Shiu G 2021 Probing the structure of string theory vacua with genetic algorithms and reinforcement learning *35th Conf. on Neural Information Processing Systems* p 11
- [34] Niarchos V, Papageorgakis C, Richmond P, Stapleton A G and Woolley M 2023 Bootstrability in line-defect CFTs with improved truncation methods *Phys. Rev. D* **108** 105027
- [35] Berglund P, He Y-H, Heyes E, Hirst E, Jejjala V and Lukas A 2024 New Calabi–Yau manifolds from genetic algorithms *Phys. Lett. B* **850** 138504
- [36] Gukov S, Halverson J and Ruehle F 2024 Rigor with machine learning from field theory to the Poincaré conjecture (arXiv:2402.13321 [hep-th])
- [37] Ernst D and Louette A 2024 *Introduction to Reinforcement Learning* ed S Feuerriegel, J Hartmann, C Janiesch and P Zschech pp 111–26
- [38] He Y H 2017 Deep-learning the landscape (arXiv:1706.02714 [hep-th])
- [39] Carifio J, Halverson J, Krioukov D and Nelson B D 2017 Machine learning in the string landscape *J. High Energy Phys.* **JHEP09(2017)157**
- [40] Krefl D and Seong R-K 2017 Machine learning of Calabi-Yau volumes *Phys. Rev. D* **96** 066014
- [41] Ruehle F 2017 Evolving neural networks with genetic algorithms to study the String Landscape *J. High Energy Phys.* **JHEP08(2017)038**

- [42] Hughes M C 2016 A neural network approach to predicting and computing knot invariants (arXiv:[1610.05744](https://arxiv.org/abs/1610.05744))
- [43] Jejjala V, Kar A and Parrikar O 2019 Deep learning the hyperbolic volume of a knot *Phys. Lett. B* **799** 135033
- [44] Gukov S, Halverson J, Ruehle F and Sułkowski P 2021 Learning to Unknot *Mach. Learn. Sci. Tech.* **2** 025035
- [45] Craven J, Jejjala V and Kar A 2021 Disentangling a deep learned volume formula *J. High Energy Phys.* **JHEP06(2021)040**
- [46] Craven J, Hughes M, Jejjala V and Kar A 2023 Learning knot invariants across dimensions *SciPost Phys.* **14** 021
- [47] Davies A *et al* 2021 Advancing mathematics by guiding human intuition with ai *Nature* **600** 70–74
- [48] Gukov S, Halverson J, Manolescu C and Ruehle F 2023 Searching for ribbons with machine learning (arXiv:[2304.09304](https://arxiv.org/abs/2304.09304))
- [49] Putrov P and Ri S J 2023 Graph neural networks and 3-dimensional topology (arXiv:[2305.05966](https://arxiv.org/abs/2305.05966))
- [50] Casler C 1965 An embedding theorem for connected 3-manifolds with boundary *Proc. Am. Math. Soc.* **16** 559–66
- [51] Burton B A 2007 Structures of small closed non-orientable 3-manifold triangulations *J. Knot Theory Ramifi.* **16** 545–74
- [52] Lickorish W 1999 Simplicial moves on complexes and manifolds *Geometry & Topology Monographs Volume 2: Proceedings of the Kirbyfest* pp 299–320
- [53] Culler M, Dunfield N M, Goerner M and Weeks J R SnapPy, a computer program for studying the geometry and topology of 3-manifolds (available at: <http://snappy.computop.org>)
- [54] Burton B A *et al* 1999–2021 Regina: software for low-dimensional topology (available at: [HTTP://REGINA-NORMAL.GITHUB.IO/](http://REGINA-NORMAL.GITHUB.IO/))
- [55] Burton B A 2011 The Pachner graph and the simplification of 3-sphere triangulations *Proc. 27th Annual ACM Symposium on Computational Geometry - SoCG'11* (ACM Press)
- [56] Hagberg A A, Schult D A and Swart P 2008 Exploring network structure, dynamics, and function using networkx
- [57] Pedregosa F *et al* 2011 Scikit-learn: machine learning in Python *J. Mach. Learn. Res.* **12** 2825–30
- [58] Abadi M *et al* 2015 TensorFlow: large-scale machine learning on heterogeneous systems (available at: [www.tensorflow.org/](http://www.tensorflow.org/))
- [59] Matveev S 1988 Transformations of special spines and the Zeeman conjecture *Math. USSR Izvestia* **31** 423–34
- [60] Piergallini R 1988 Standard moves for standard polyhedra and spines *Rend. Circ. Mat. Palermo* **37** 391–414
- [61] Lackenby M 2020 Algorithms in 3-manifold theory (arXiv:[2002.02179](https://arxiv.org/abs/2002.02179))
- [62] Matveev S 2003 *Algorithmic Topology and Classification of 3-Manifolds (Algorithms and Computation in Mathematics)* vol 9 (Springer)
- [63] Kuperberg G 2019 Algorithmic homeomorphism of 3-manifolds as a corollary of geometrization *Pac. J. Math.* **301** 189–241
- [64] Scott P and Short H 2014 The homeomorphism problem for closed 3-manifolds *Algebr. Geom. Topol.* **14** 2431–44
- [65] Rubinstein H 1995 An algorithm to recognize the three sphere *Proc. Int. Congress of Mathematicians (Zurich 1994)* (Birkhäuser) pp 601–11
- [66] Mijatović A 2003 Simplifying triangulations of  $S^3$  *Pac. J. Math.* **208** 291–324
- [67] Chicco D and Jurman G 2020 The advantages of the matthews correlation coefficient (mcc) over f1 score and accuracy in binary classification evaluation *BMC Genomics* **21** 6
- [68] King T, Butcher S and Zalewski L 2017 Apocrita - high performance computing cluster for queen mary University of London (<https://doi.org/10.5281/zenodo.438045>)
- [69] Holland P W and Leinhardt S 1971 Transitivity in structural models of small groups *Comp. Group Stud.* **2** 107–24
- [70] Watts D J and Strogatz S 1998 Collective dynamics of ‘small-world’ networks *Nature* **393** 440–2
- [71] Wiener H 1947 Structural determination of paraffin boiling points *J. Am. Chem. Soc.* **69** 17–20
- [72] Perron O 1907 Zur theorie der matrices *Math. Ann.* **64** 248–63
- [73] Frobenius G 1912 Ueber matrizen aus nicht negativen elementen *Sitzungsber. Koni. Preuss. Akad. Wiss.* 456–77
- [74] Bonacich P 1987 Power and centrality: A family of measures *Am. J. Sociol.* **92** 1170–82
- [75] Liebchen C and Rizzi R 2007 Classes of cycle bases *Discrete Appl. Math.* **155** 337–55

- [76] Gurney K 2018 *An Introduction to Neural Networks* (CRC Press)
- [77] Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez A N, Kaiser L and Polosukhin I 2023 Attention is all you need
- [78] Lin T, Wang Y, Liu X and Qiu X 2022 A survey of transformers *AI Open* **3** 111–32
- [79] OpenAI 2023 Gpt-4 technical report (arXiv:[2303.08774](#))
- [80] Douglas M R 2023 Large language models (arXiv:[2307.05782](#) [cs.CL])
- [81] Devlin J, Chang M W, Lee K and Toutanova K 2019 Bert: Pre-training of deep bidirectional transformers for language understanding
- [82] Wolf T *et al* 2020 Transformers: state-of-the-art natural language processing *Proc. 2020 Conf. on Empirical Methods in Natural Language Processing: System Demonstrations* (Association for Computational Linguistics) pp 38–45
- [83] Gordon C M and Luecke J 1989 Knots are determined by their complements *Bull. Am. Math. Soc* **20** 83–87
- [84] Dehn M 1938 Die Gruppe der Abbildungsklassen: Das arithmetische Feld auf Flächen *Acta Math.* **69** 135–206
- [85] Burton B A 2011 Simplification paths in the pachner graphs of closed orientable 3-manifold triangulations (arXiv:[1110.6080](#))
- [86] ANSI 1975 ISO-IR-6: ASCII Graphic character set