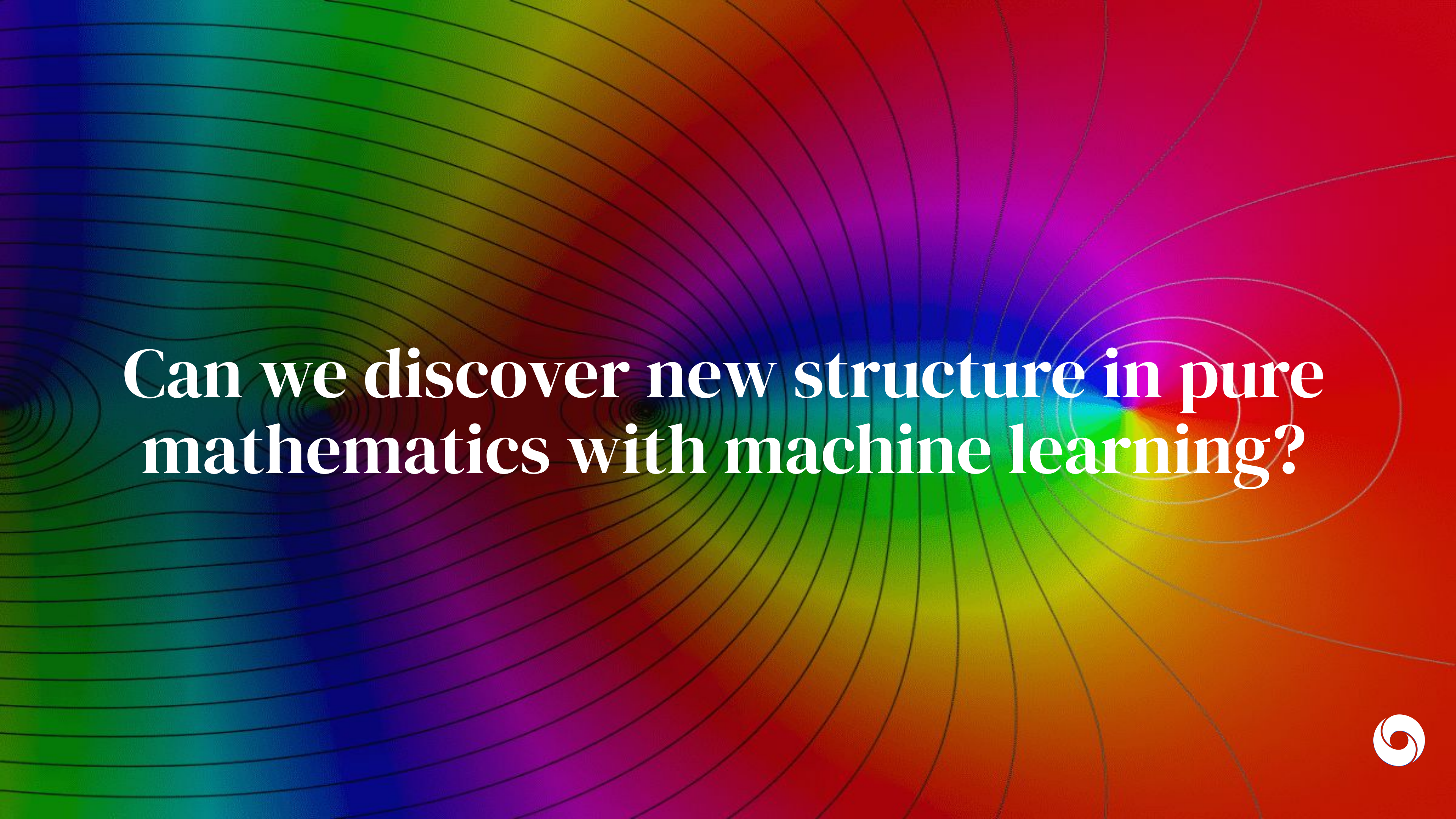
DeepMind

# AI x Mathematics

Petar Veličković

AI + Science @ NeurIPS 2021
13 December 2021

with Alex Davies, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Charles Blundell, Peter Battaglia, András Juhász (Oxford), Marc Lackenby (Oxford), Geordie Williamson (Sydney), Demis Hassabis and Pushmeet Kohli

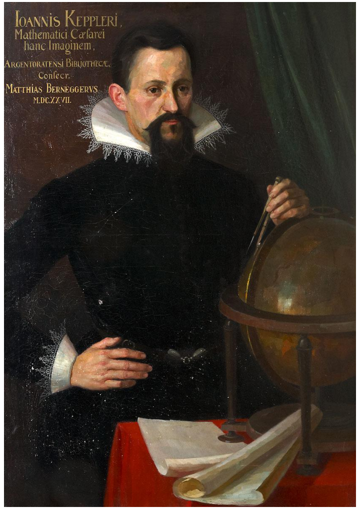


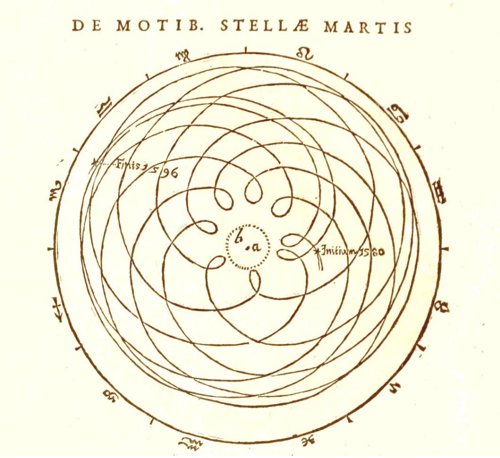
# Can we discover new structure in pure mathematics with machine learning?

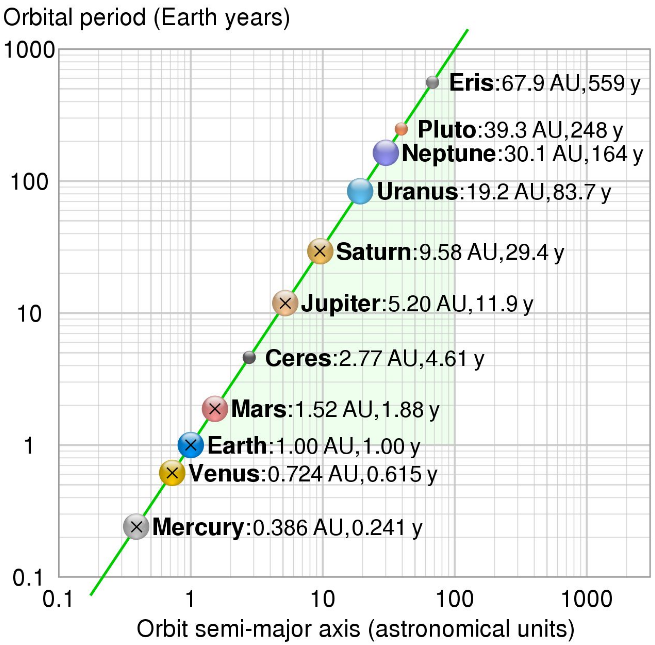
# Srinivasa Ramanujan

> "The great advances in mathematics have not been made by logic but by creative imagination
>
> – *George Frederick James Temple* on Ramanujan

# Kepler’s laws of planetary motion

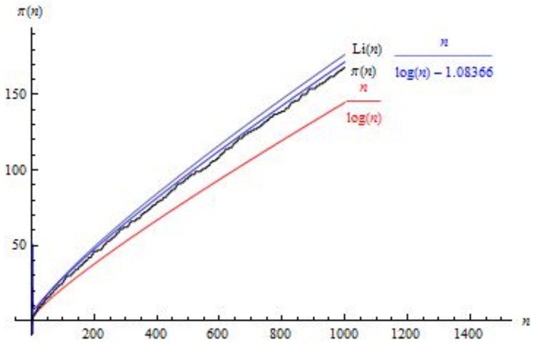




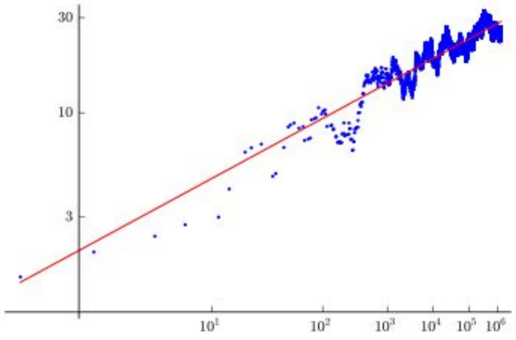

*“I first believed I was dreaming… But it is absolutely certain and exact that the ratio which exists between the period times of any two planets is precisely the ratio of the 3/2th power of the mean distance.”*

# Mathematical pattern discovery



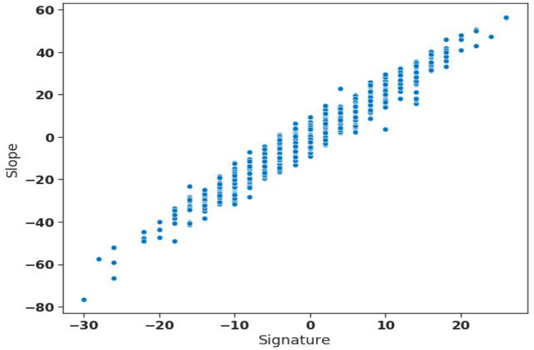

**Prime number theorem**
Legendre, 1800

1. Manually calculated examples
2. Manually detected pattern




**BSD conjecture**
Birch and Swinnerton-Dyer, 1965

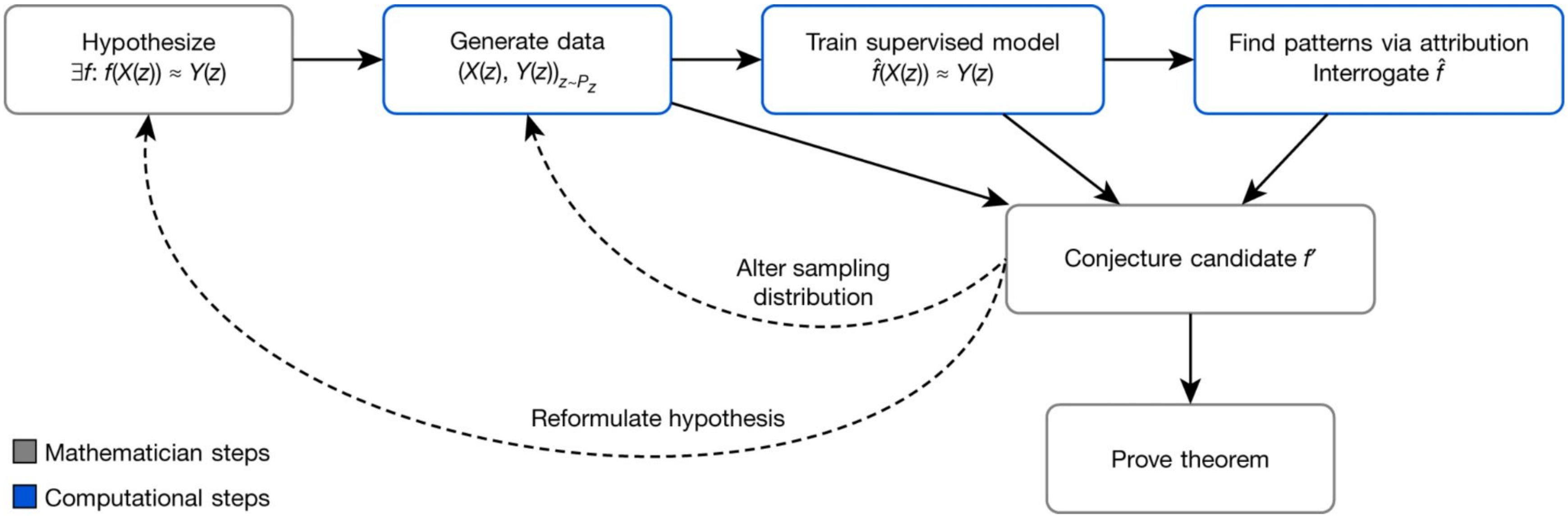
1. Computed examples
2. Manually detected pattern

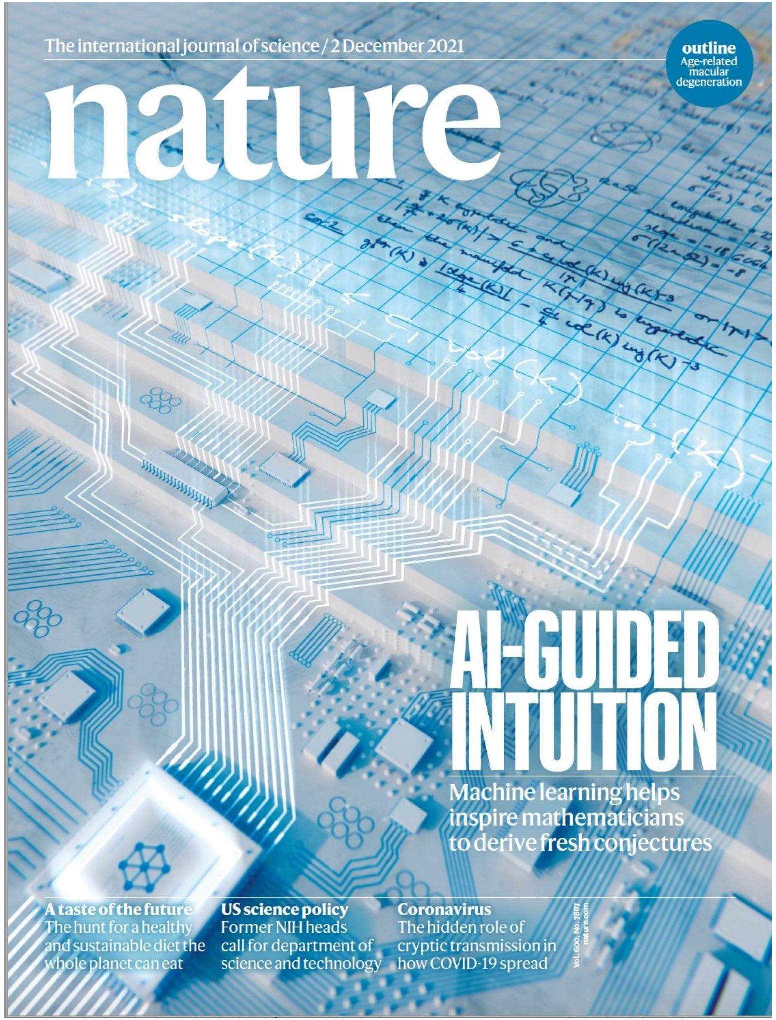



**Our work**
DeepMind, 2021

1. Computed examples
2. Computationally detected patterns

# AI-assisted mathematical discovery






"This paper marks the beginning of a new phase in the use of computers in mathematical research."

"I cannot imagine such a mathematician not using these methods, where available, given that not using them means progressing much more slowly and being unable to analyze very complicated objects and very large datasets."

"Such a new approach (even conjectural) to the combinatorial invariance conjecture would certainly be publishable in a top combinatorics journal."

"I judge the separate mathematical paper "The signature and cusp geometry of hyperbolic knots" likely to be accepted by the top speciality journal in the field (Geometry and Topology) and be a shoo in at the second best one (J. Topology)."

# Knot Theory

## THE SIGNATURE AND CUSP GEOMETRY OF HYPERBOLIC KNOTS

ALEX DAVIES, ANDRÁS JUHÁSZ, MARC LACKENBY, AND NENAD TOMASEV

ABSTRACT. We introduce a new real-valued invariant called the natural slope of a hyperbolic knot in the 3-sphere, which is defined in terms of its cusp geometry. We show that twice the knot signature and the natural slope differ by at most a constant times the hyperbolic volume divided by the cube of the injectivity radius. This inequality was discovered using machine learning to detect relationships between various knot invariants. It has applications to Dehn surgery and to 4-ball genus. We also show a refined version of the inequality where the upper bound is a linear function of the volume, and the slope is corrected by terms corresponding to short geodesics that link the knot an odd number of times.

https://arxiv.org/abs/2111.15323


Marc Lackenby & András Juhász

UNIVERSITY OF OXFORD

# Representation Theory

## TOWARDS COMBINATORIAL INVARIANCE FOR KAZHDAN-LUSZTIG POLYNOMIALS

CHARLES BLUNDELL, LARS BUESING, ALEX DAVIES, PETAR VELIČKOVIĆ, AND GEORDIE WILLIAMSON

ABSTRACT. Kazhdan-Lusztig polynomials are important and mysterious objects in representation theory. Here we present a new formula for their computation for symmetric groups based on the Bruhat graph. Our approach suggests a solution to the combinatorial invariance conjecture for symmetric groups, a well-known conjecture formulated by Lusztig and Dyer in the 1980s.

https://arxiv.org/abs/2111.15161

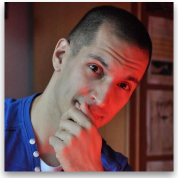
Geordie Williamson

THE UNIVERSITY OF SYDNEY

DeepMind

# Knot Theory

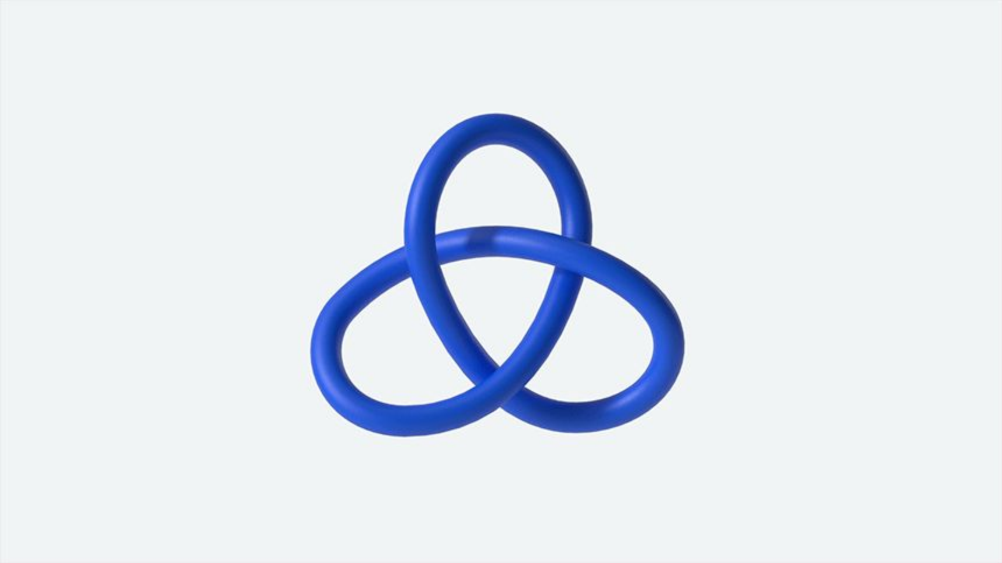


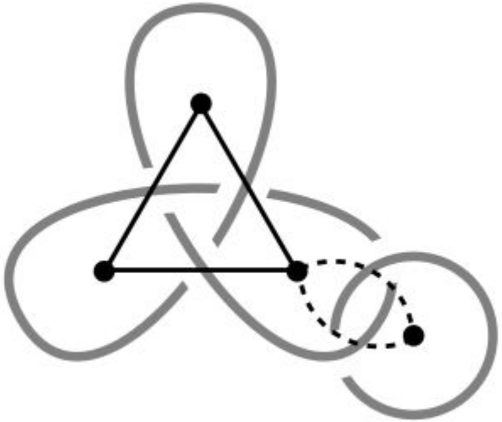
**The signature and cusp geometry of hyperbolic knots.**
Davies, Juhász, Lackenby, Tomašev. *arXiv:2111.15323*

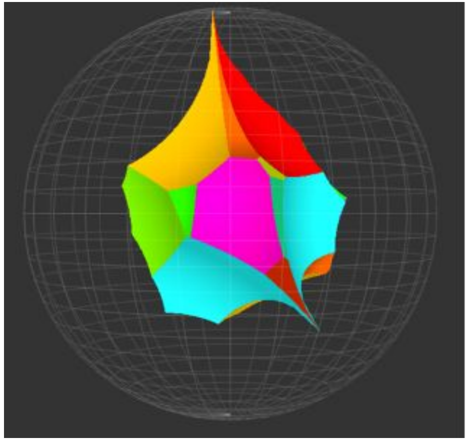
# Knots

A closed loop embedded in 3 dimensions.



# Knot invariants

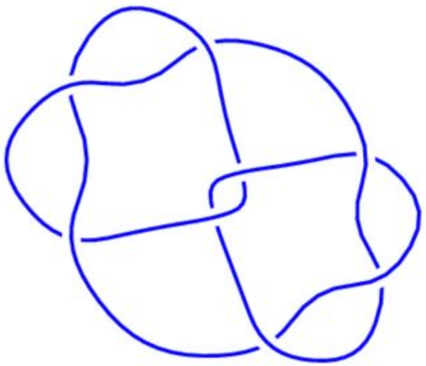




There are few known connections between the geometry and algebraic structure of knots.

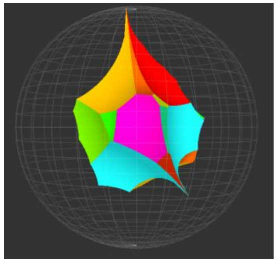
# Hypothesis

Is there a **relationship** between the **geometry** and **algebraic structure** of a knot?

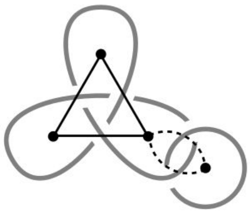


**Knot**

**3D Geometric invariants**



**4D Algebraic/Quantum invariants**



# Example knots and invariants

| $z$: **Knot** | $X(z)$: **Geometric Invariants** | | | | $Y(z)$: **Algebraic Invariants** | | |
|---|---|---|---|---|---|---|---|
| | Volume | Chern–Simons | Meridional Translation | ... | Signature | Jones Polynomial | ... |
|  | 2.0299 | 0 | $i$ | ... | 0 | $t^{-2} - t^{-1} + 1 - t + t^2$ | ... |
|  | 2.8281 | -0.1532 | $.7381 + 0.8831i$ | ... | -2 | $t - t^2 + 2t^3 - t^4 + t^5 - t^6$ | ... |
|  | 3.1640 | 0.1560 | $-.7237 + 1.0160i$ | ... | 0 | $t^{-2} - t^{-1} + 2 - 2t + t^2 - t^3 + t^4$ | ... |

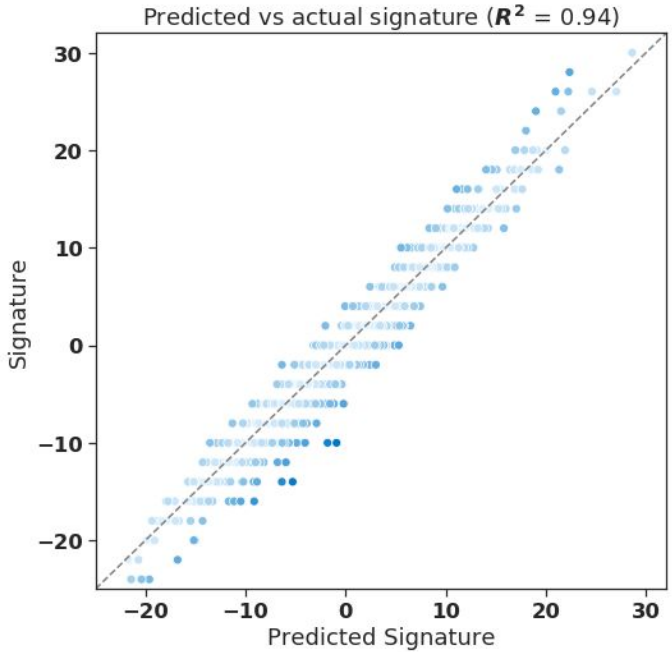
# Example knots and invariants

| $z$: **Knot** | $X(z)$: **Geometric Invariants** | | | | $Y(z)$: **Algebraic Invariants** | | |
|---|---|---|---|---|---|---|---|
| | Volume | Chern–Simons | Meridional Translation | … | Signature | Jones Polynomial | … |
|  | 2.0299 | 0 | $i$ | … | 0 | $t^{-2} - t^{-1} + 1 - t + t^2$ | … |
|  | 2.8281 | -0.1532 | $.7381 + 0.8831i$ | … | -2 | $t - t^2 + 2t^3 - t^4 + t^5 - t^6$ | … |
|  | 3.1640 | 0.1560 | $-.7237 + 1.0160i$ | … | 0 | $t^{-2} - t^{-1} + 2 - 2t + t^2 - t^3 + t^4$ | … |

Train an **MLP** to predict the **signature** (Y) from the **geometric invariants** (X)

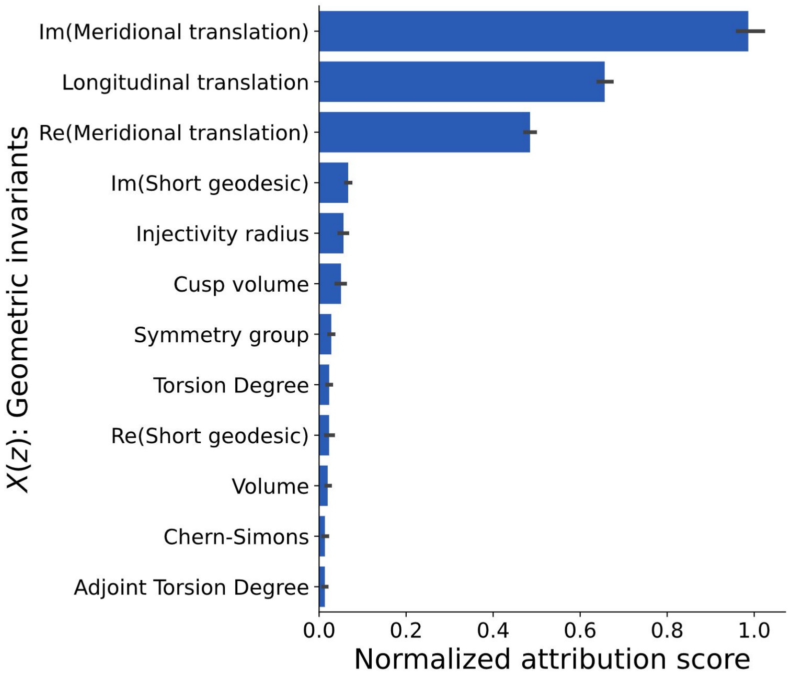
# Training the model

The signature (algebraic structure) can be **well predicted** from the geometric structure.



# Interrogating the model

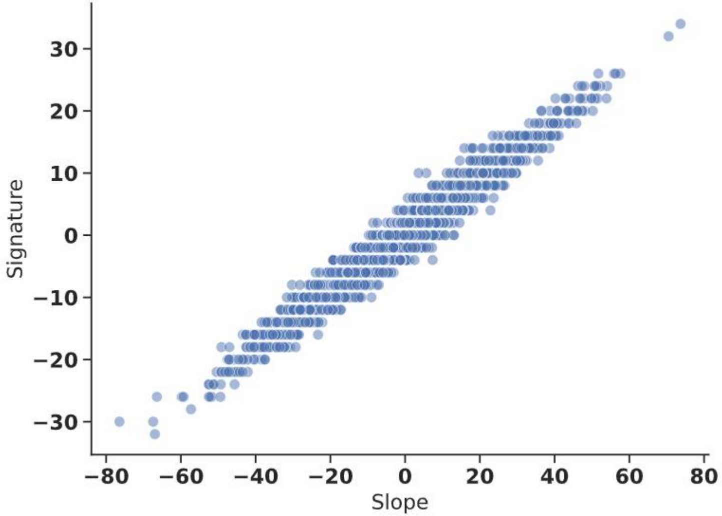
Analysing feature saliency gives us the **relevant** parts of the geometry for predicting the signature.



# Discovering the ***natural slope***

**Meridional** translation ($M(K)$)
**Longitudinal** translation ($L(K)$)

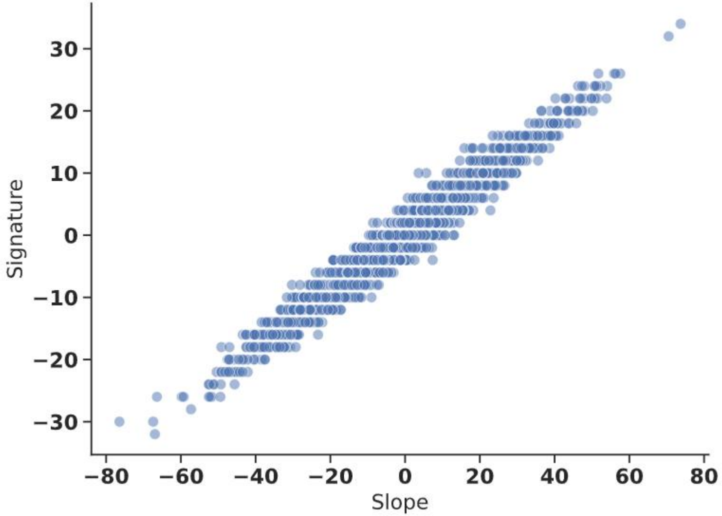
$$\text{slope}(K) = \frac{\text{Re}(M(K)) + \text{Re}(L(K))}{\|M(K)\|^2}$$



# The “signature-slope” conjecture

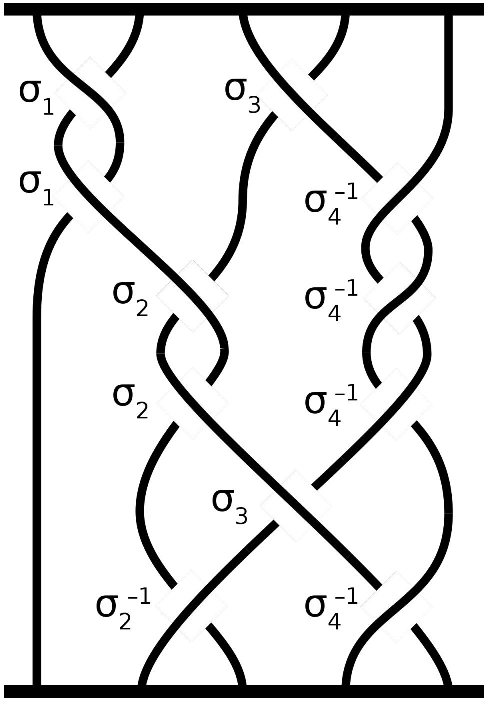
Possible relationship **obvious** to spot
Empirically validated over many knots

$$|2\sigma(K) - \mathrm{slope}(K)| < \sqrt{\mathrm{vol}(K)} + c$$



# ...but wait!

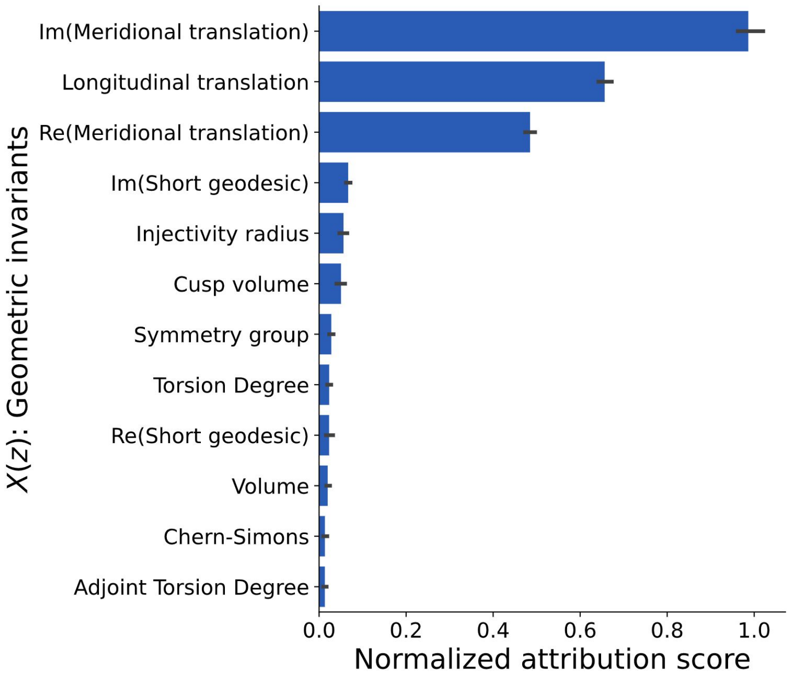
Can construct a **counterexample** using *braided knots*



# Interrogating the model *again*!

Maybe we can make use of some **more** geometric invariants?

Short geodesic
Injectivity radius
Cusp volume



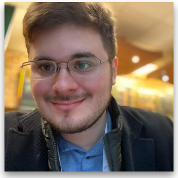
# The theorem

$$|2\sigma(K) - \mathrm{slope}(K)| \leq c_3 \mathrm{vol}(K) \mathrm{inj}(K)^{-3}.$$

DeepMind

# Representation Theory




**Towards combinatorial invariance for Kazhdan–Lusztig polynomials.**
Blundell, Buesing, Davies, Veličković, Williamson. *arXiv: 2111.15161*

# Where do we even begin?

What I will present to you is a **streamlined** view of 1.5 years of research.

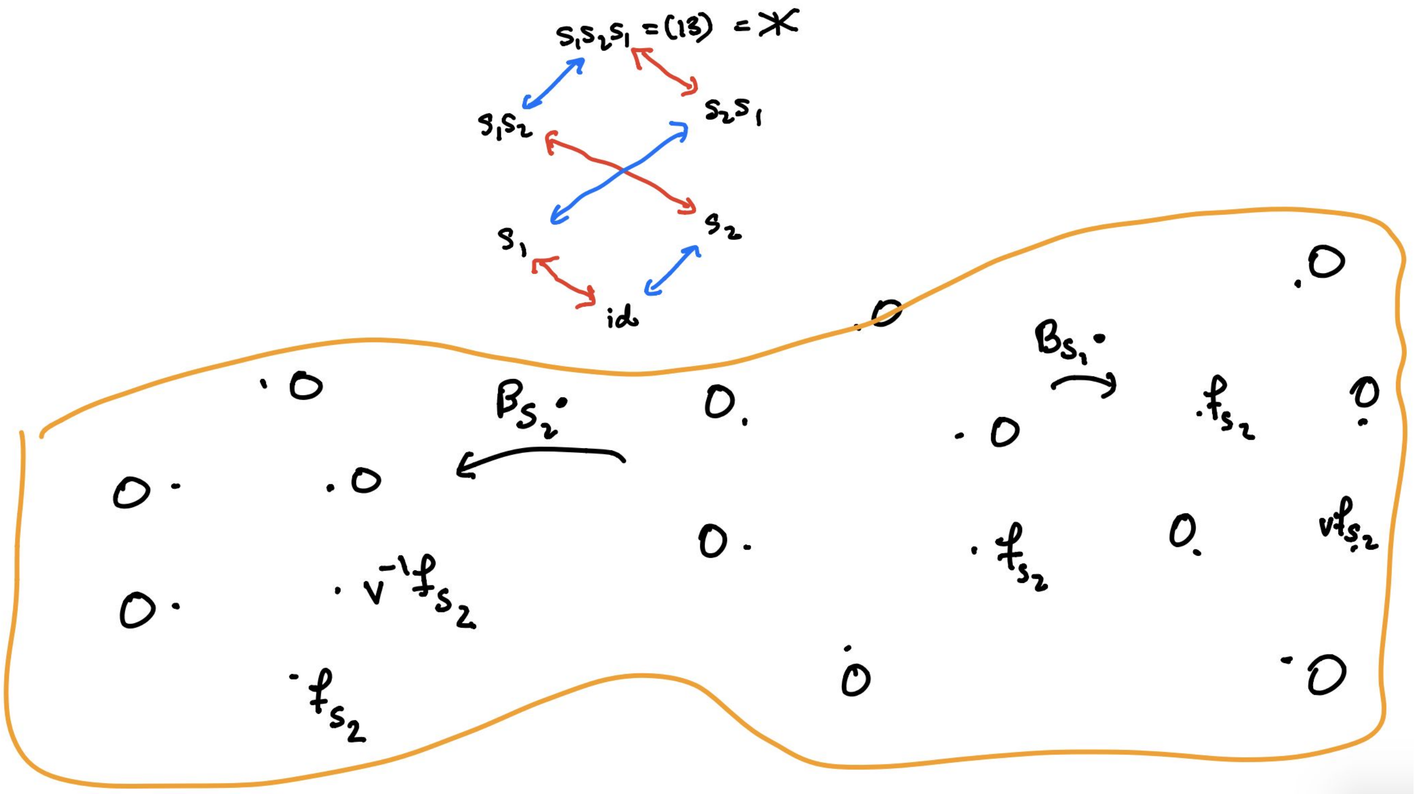
- I condense only the theory necessary to have a basic understanding of the problem

At the onset, we knew next-to-nothing about representation theory, and Geordie wasn't very familiar with machine learning.

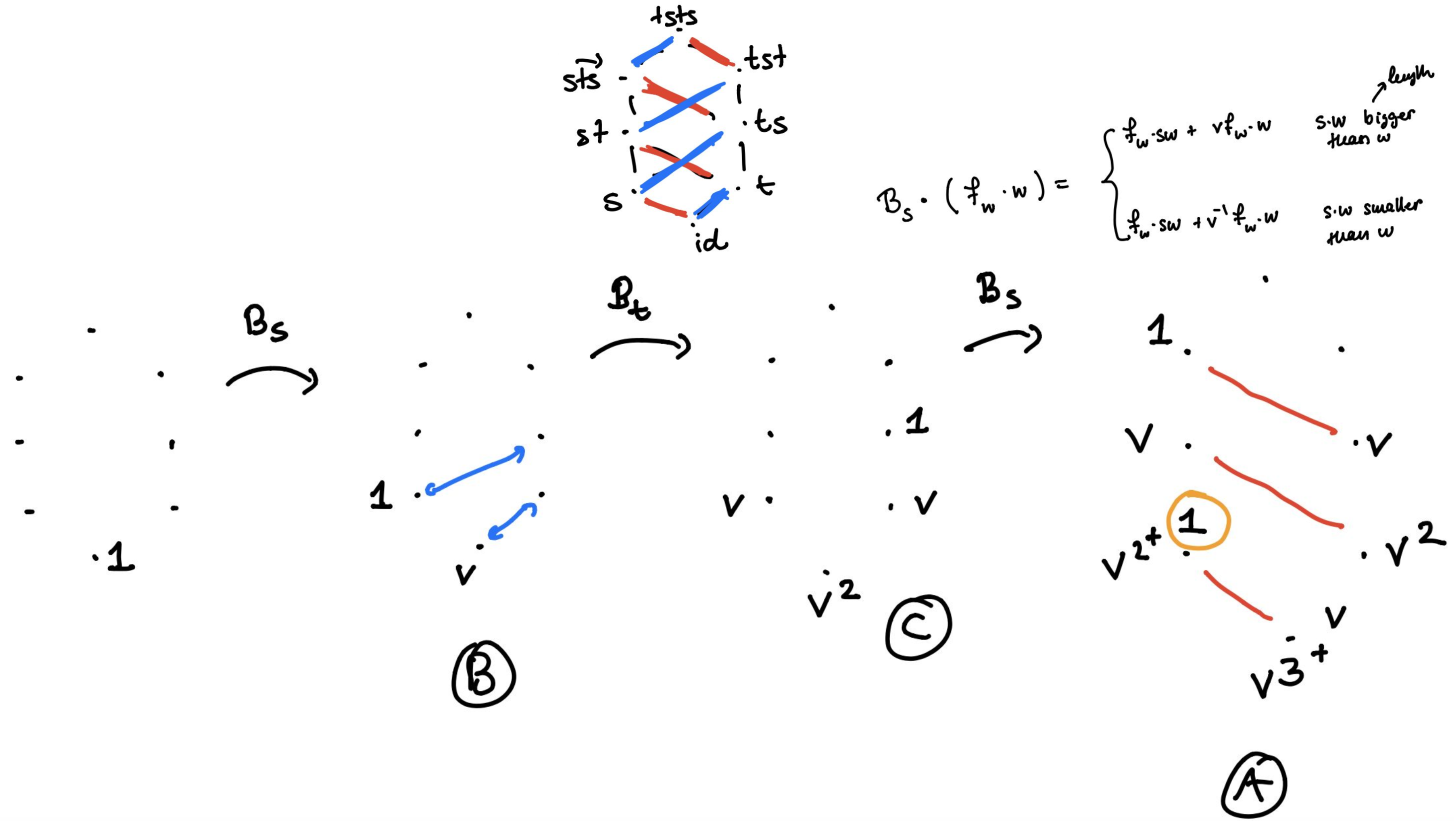
Many of our weekly sessions were 2h lectures and "deep dives" to get us all on the same page.

**BUT**, most of the intermediate machine learning work we've done was impactful to the final result.

# Snapshots from representation theory lectures

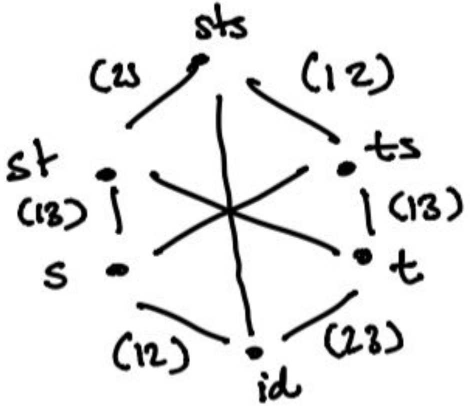


# Snapshots from representation theory lectures



First we define another class of polynomials, "R-polynomials".

Need: ① Bruhat graph of our interval.
<u>WITH</u> labels

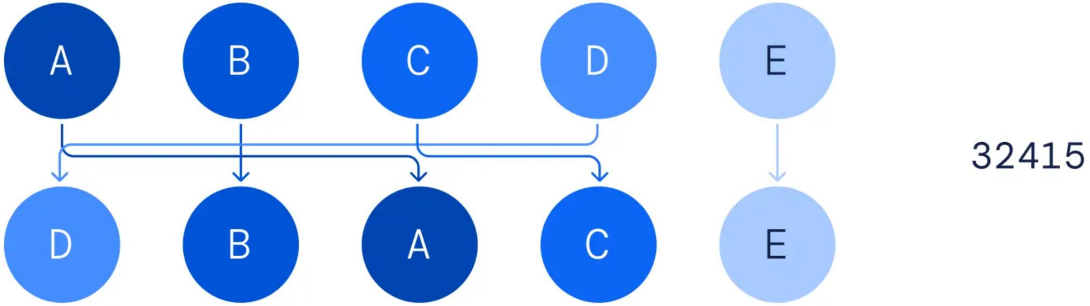



② Reflection ordering: ordering on reflections satisfying ...

USE LEXICOGRAPHIC : $(12) < (13) < (23)$

# Symmetry groups

Groups of **transformations** that leave the underlying object **unchanged**




Here we focus on ***permutation groups*** (the essence behind graph neural networks!)

# Coxeter groups

**Coxeter groups** can be fully described in terms of a subset of the transformations (*"reflections"*)

For the permutation group, these reflections are single **swaps**!

In the 4-element permutation group ($S_4$), there are **six** reflections

**2134, 3214, 4231, 1324, 1432, 1243**

# Generators of the group

**Generators**: Set of symmetries from which you can recover the **entire group** by **composition**

For permutation groups, the generators are swaps of **adjacent elements**

In the 4-element permutation group ($S_4$), there are **three** simple reflections

**2134**, 3214, 4231, **1324**, 1432, **1243**

# Coxeter groups as *directed graphs*

We can think of each permutation as a **node** in a graph (**labelled Bruhat graph**)!

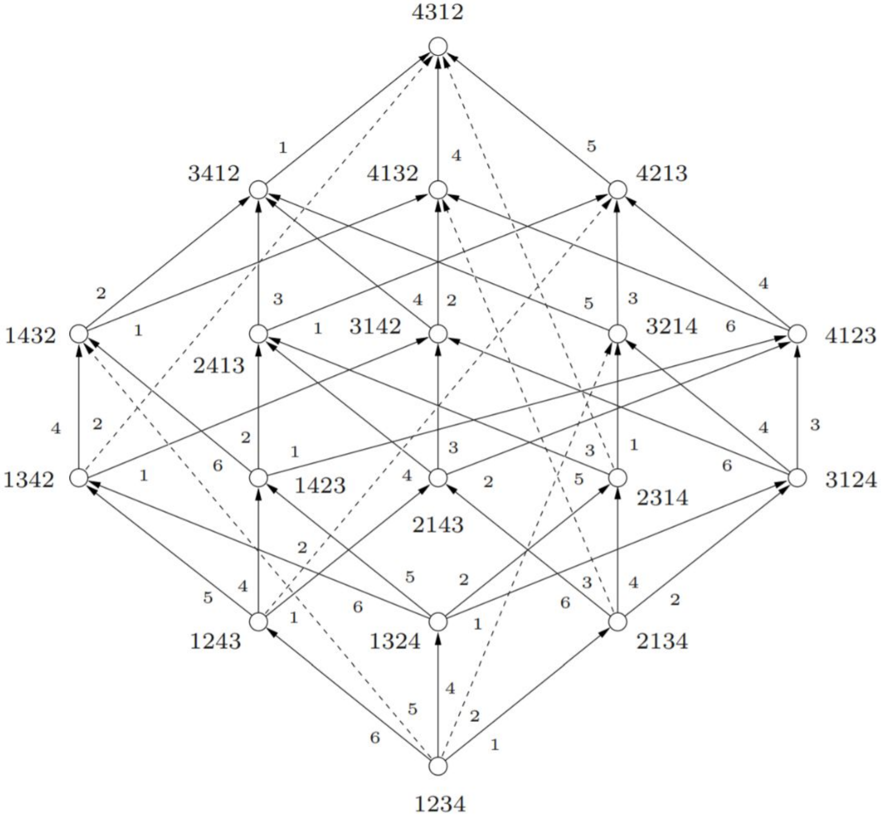
Draw an edge between two nodes if you can recover one from the other by a **reflection** (swap)

Orient the graph as follows:

- The identity permutation (1234 for $S_4$) at the bottom
- On the first level, all permutations reachable from identity by a simple reflection
    - Adjacent swaps!
- Proceed inductively

Each edge should be directed towards the “higher” permutation in the graph.

# Labelled Bruhat graph for $S_4$



# Combinatorial invariance conjecture

Fix any two elements in a Coxeter group (e.g. two permutations in $S_n$).

From these two elements, you can define *two* objects:

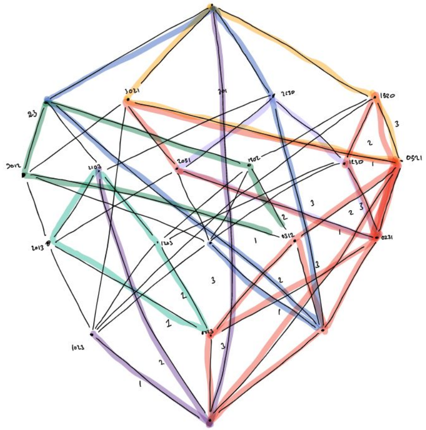
- **Bruhat interval**: *subgraph* of the **unlabelled** Bruhat graph containing all paths between them
- **Kazhdan–Lusztig polynomial**: a polynomial with *integer* coefficients

**NB** the Bruhat interval is ***unlabelled***: we don't know which node/edge belongs to which permutation!

It has been long conjectured (by Lusztig and Dyer), that these two objects are fundamentally related.
(*i.e.*, that we can compute the KL polynomial from the Bruhat interval)

This is the **combinatorial invariance** conjecture, which has stood for 50 years!

# **Combinatorial invariance** conjecture

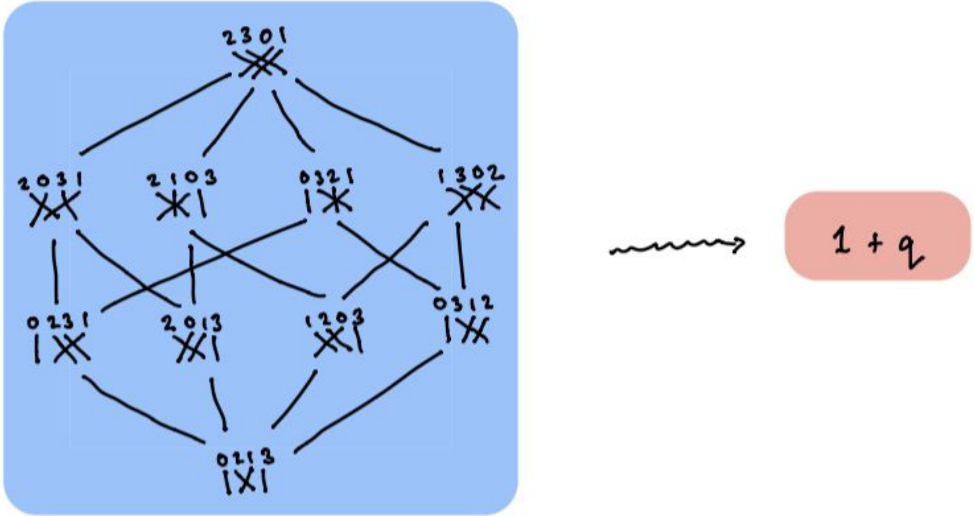


$1 + q$

Is there a map from a **Bruhat interval** to its associated **KL polynomial**?

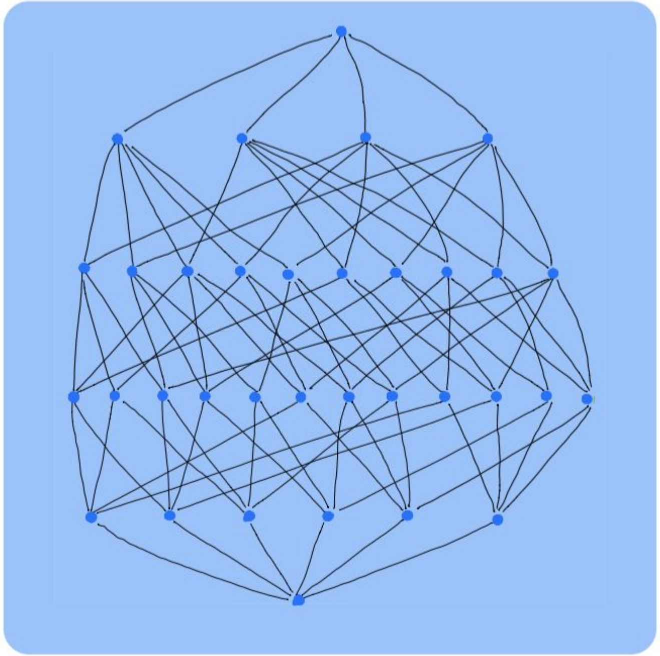
# Patterns emerge quite late!

For **most** intervals we can easily compute, the KL polynomial is very simple! (e.g. 1, q or 1 + q)!



# Patterns emerge quite late!

For **most** intervals we can easily compute, the KL polynomial is very simple! (e.g. 1, q or 1 + q)!

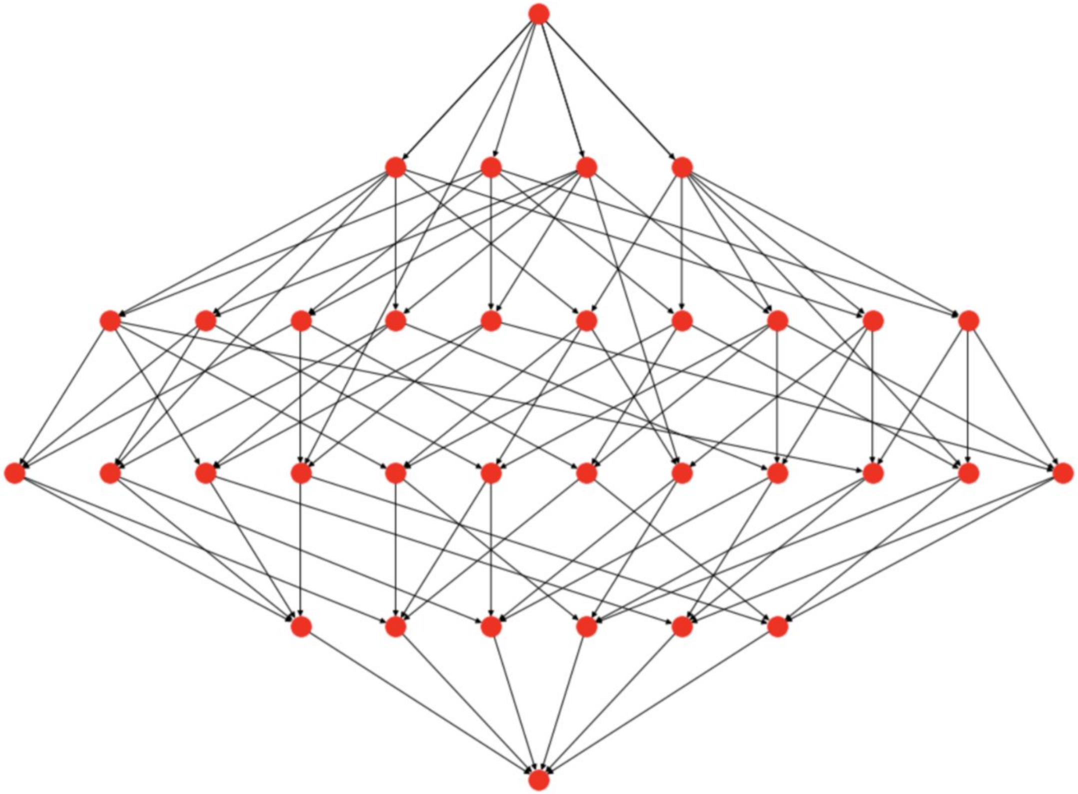


$\rightsquigarrow$ $\boxed{1+q^2}$

Remark: This is one of the easiest examples.

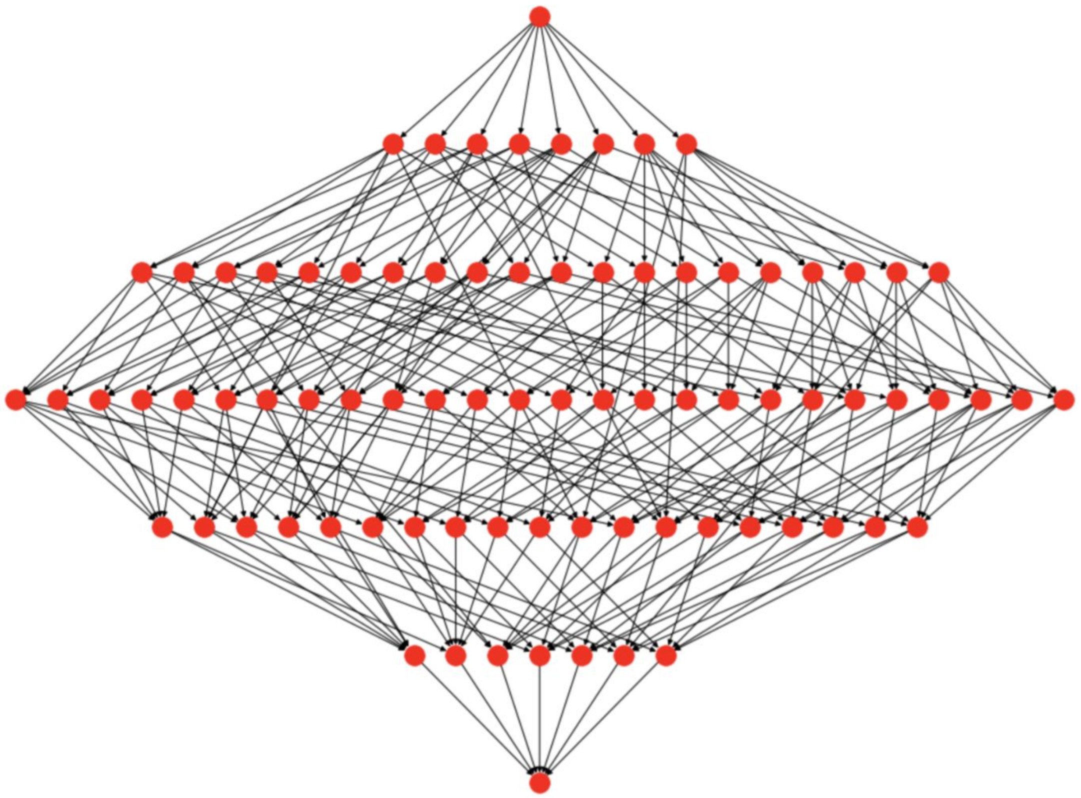
Moral: Bruhat intervals become complicated and difficult to visualize very quickly!

# Training data for the model ($S_6$ example)



$\mapsto 1+q^2$

# Training data for the model ($S_6$ example)



$\mapsto 1+3q+q^2$

# Which model to use?

We studied known algorithms which computed the KL polynomial from the **labelled** Bruhat graph

It felt a lot like ***message passing*** with **ReLU** activation!

$\Rightarrow$ we employ *graph neural networks* (**GNNs**)!

*Graph classification* tasks

(predict coefficients of the KL polynomial)

Hypothesis: the GNN would be able to recover the relevant parts of these equations from the structural properties of the Bruhat graph alone

$$h_{y,x} = \tau_{>0}\left( \sum_{y<z\leq x} r_{y,z} \overline{h_{z,x}} \right)$$

throw away all coefficients of $1, v^{-1}, v^{-2}, \dots$

$\overline{\phantom{h}}$ : replace $v$ by $v^{-1}$.

$$r_{y,x} = \sum_{\substack{\text{increasing paths from} \\ y \to x \text{ in} \\ \text{Bruhat graph}}} (v - v^{-1})^{\ell(\text{path})}$$

$\ell(\text{path})$

E.g.: $r_{id,\, sts} = (v-v^{-1})^3 + (v-v^{-1})$

# Training the GNN

Build a dataset of intervals up to $S_9$

After training the GNN, it learns to predict KL coefficients significantly better than chance!

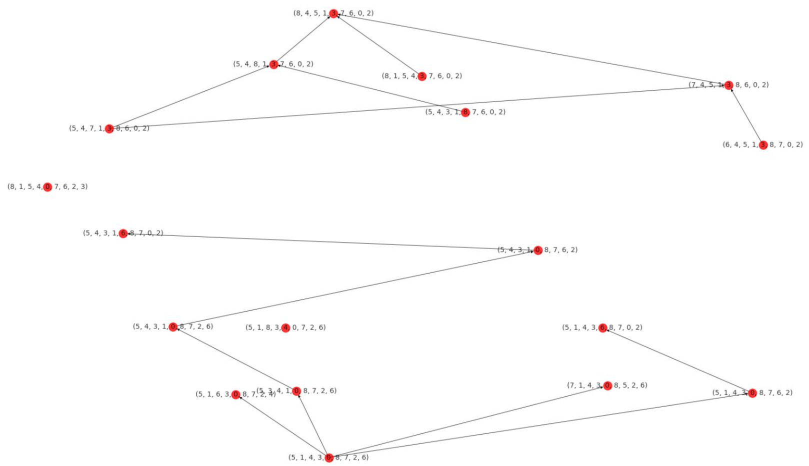
| | $q$ | $q^2$ | $q^3$ | $q^4$ |
|---|---|---|---|---|
| Baseline accuracy | 21% | 12% | 29% | 88% |
| Full interval test accuracy | 98% | 63% | 72% | 98% |

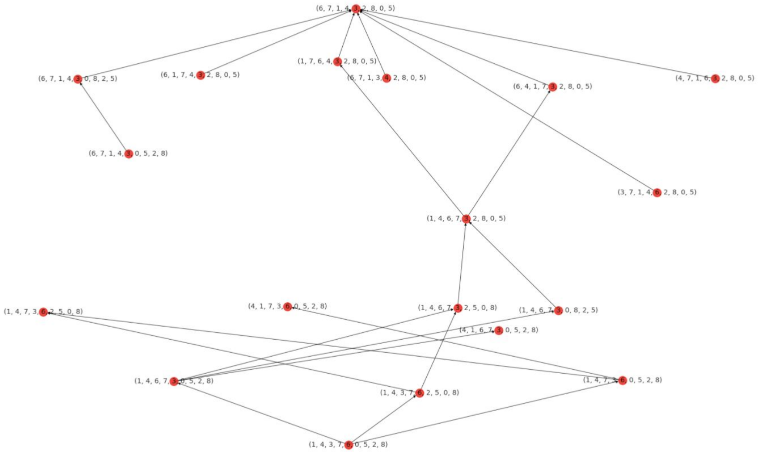
An encouraging result – previously, an algorithm for computing q was known, and little else.

What did the GNN learn to do?

# Interrogating the GNN

Use gradient saliency to identify salient nodes and edges by the GNN

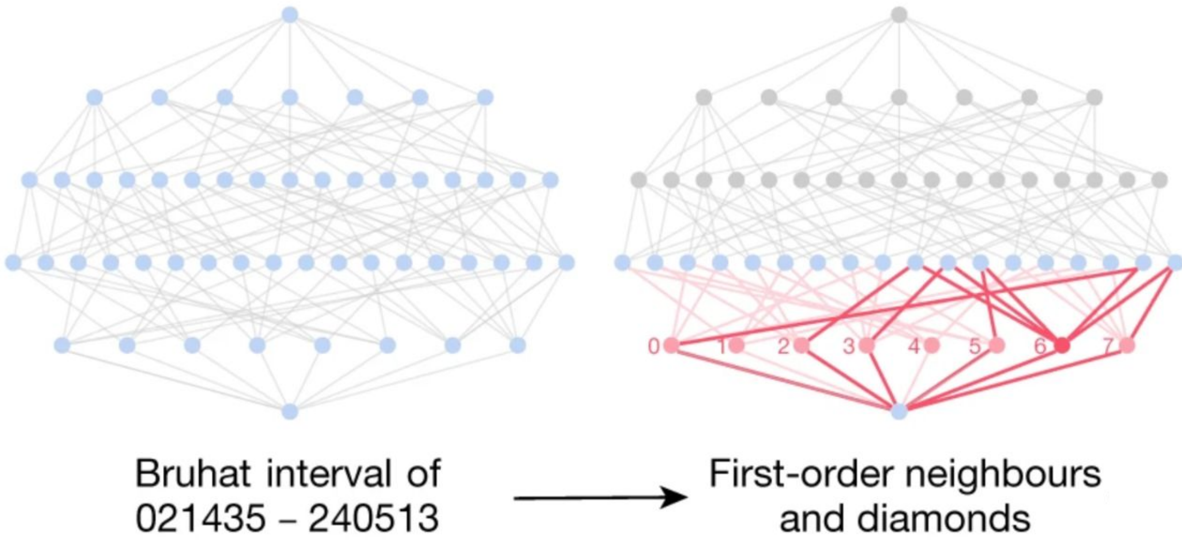


# Dihedral interval discovery

To us, the salient subgraphs made little sense...

Our collaborator realised they often featured nodes near the bottom of the interval, and related the salient subgraph to a previously studied structure: the *dihedral intervals* and *diamonds*



# Retraining the GNN

Now, train the GNN with these additional structural annotations...

| | $q$ | $q^2$ | $q^3$ | $q^4$ |
|---|---|---|---|---|
| Baseline accuracy | 21% | 12% | 29% | 88% |
| Full interval test accuracy | 98% | 63% | 72% | 98% |

# Retraining the GNN

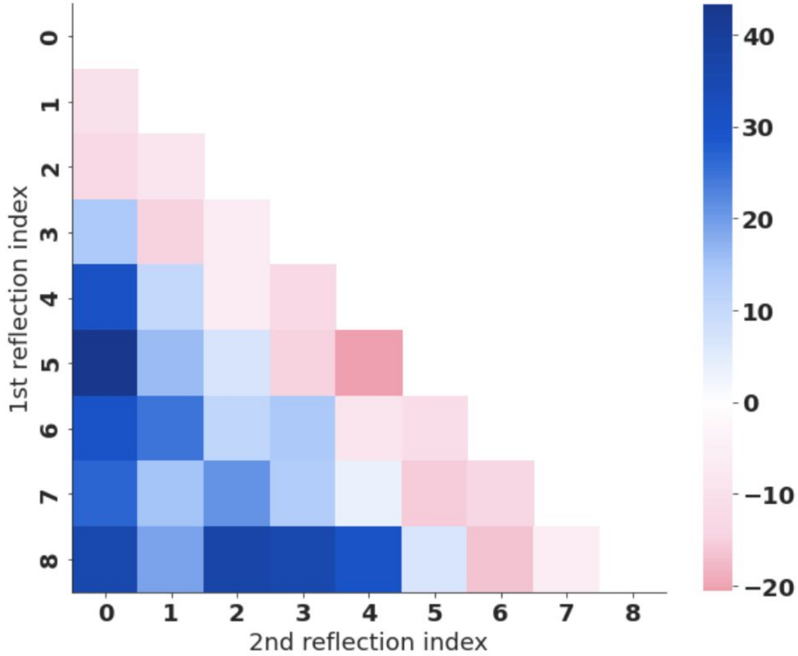
Now, train the GNN with these additional structural annotations…

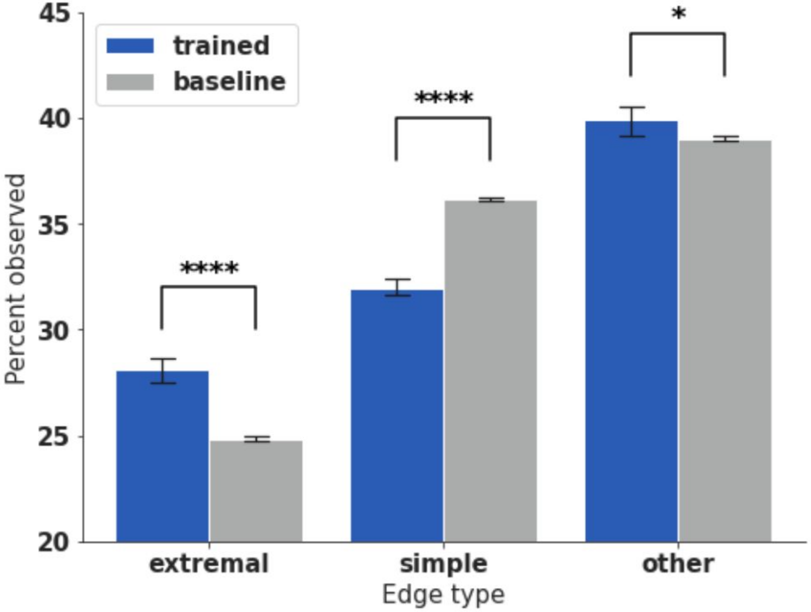
| | $q$ | $q^2$ | $q^3$ | $q^4$ |
|---|---|---|---|---|
| Baseline accuracy | 21% | 12% | 29% | 88% |
| Full interval test accuracy | 98% | 63% | 72% | 98% |
| Dihedral annotated test accuracy | 99.9% | 96.5% | 95.6% | 99.4% |

An even more encouraging result! :)

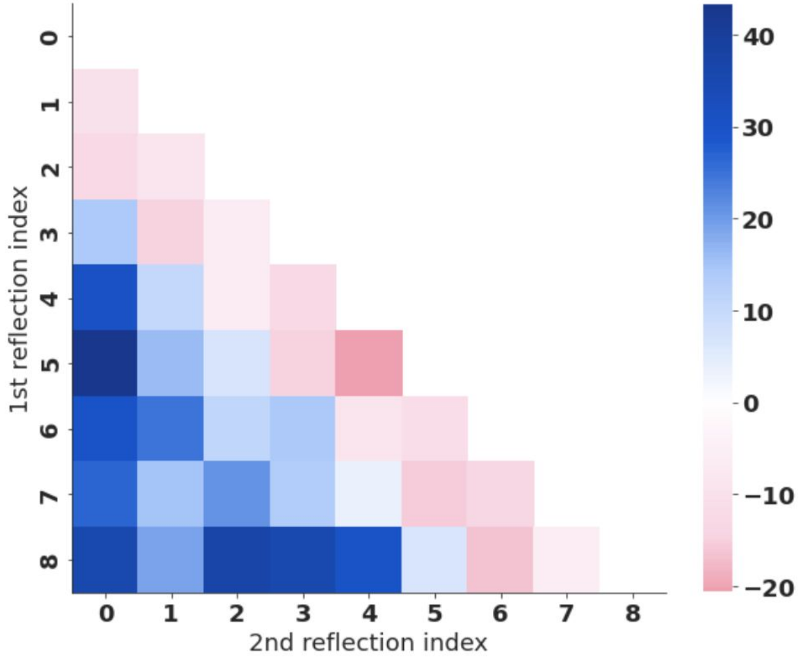
# Interrogating the new GNN

Further, we made another crucial observation: the most salient edges were ***extremal*** reflections



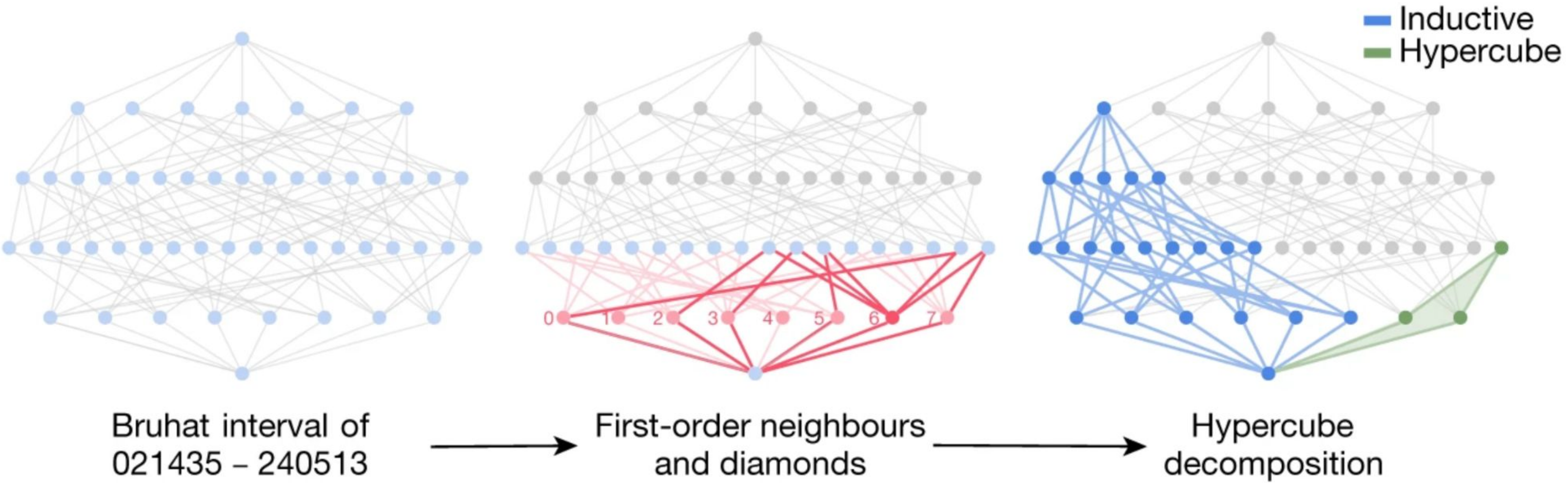
# Significance of extremal reflections




“In particular, the feeling that extremal reflections seem to play a crucial role [...] is one that I would have expected only from a handful of great world experts in representation theory, and is such a "human" suggestion that it gives me goosebumps.”

# Surprising structure

Combining these two discovered structures, Geordie proposed a ***hypercube decomposition*** of the Bruhat interval, which felt very promising for inductively computing the KL coefficients!



# Intermediate result

**Theorem**: Every Bruhat interval admits a canonical *hypercube decomposition* along its extremal reflections, from which the KL polynomial is directly computable.

# Final conjecture (proof pending!)

Remarkably, further tests suggested that *all* hypercube decompositions correctly determine the KL polynomial. This has been computationally verified for all of the ~$3*10^6$ intervals in the symmetric groups up to $S_7$ and more than $1.3*10^5$ non-isomorphic intervals sampled from the symmetric groups $S_8$ and $S_9$.

**Conjecture**: The KL polynomial of an unlabelled Bruhat interval can be calculated using the previous formula with any hypercube decomposition.

DeepMind

# Takeaways

# Takeaways

- Simple machine learning interventions can have a large impact in mathematics
  - Mathematicians are not (just) theorem provers

- When is the method particularly appropriate?

- Which areas of machine learning could (would need to?) improve
  - Explainability
  - Geometric Deep Learning

- Reach out to a mathematician near you!
  - **`https://github.com/deepmind/mathematics_conjectures`**

# Lastly…

The account I gave today is largely from a perspective of a **machine learning** researcher

If you are a mathematician, or would like to know more about the **mathematicians**' view:

THE CONVERSATION
Academic rigour, journalistic flair

## Mathematical discoveries take intuition and creativity – and now a little help from AI

December 1, 2021 5.10pm GMT

shutterstock.

Email
Twitter 26
Facebook
LinkedIn
Print

Research in mathematics is a deeply imaginative and intuitive process. This might come as a surprise for those who are still recovering from high-school algebra.

What does the world look like at the quantum scale? What shape would our universe take if we were as large as a galaxy? What would it be like to live in six or even 60 dimensions? These are the problems that mathematicians and physicists are grappling with every day.

**Author**

**Geordie Williamson**
Professor of Mathematics, University of Sydney

**Disclosure statement**

Geordie Williamson is a Professor at the University of Sydney, and a consultant in Pure Mathematics for DeepMind, a subsidiary of Alphabet.

DeepMind

# Thank you!

petarv@deepmind.com | https://petar-v.com

In collaboration with Alex Davies, Lars Buesing, Sam Blackwell, Daniel Zheng, Nenad Tomašev, Richard Tanburn, Charles Blundell, Peter Battaglia, András Juhász (Oxford), Marc Lackenby (Oxford), Geordie Williamson (Sydney), Demis Hassabis and Pushmeet Kohli