AI in Mathematics
Lecture 5
Deep Learning in Mathematics

Bar-Ilan University
Nebius Academy | Stevens Institute of Technology

April 22, 2025

About This Course

~~1 week: Intro~~

~~2 weeks: Classic ML~~

2 weeks: Deep Learning in Mathematics

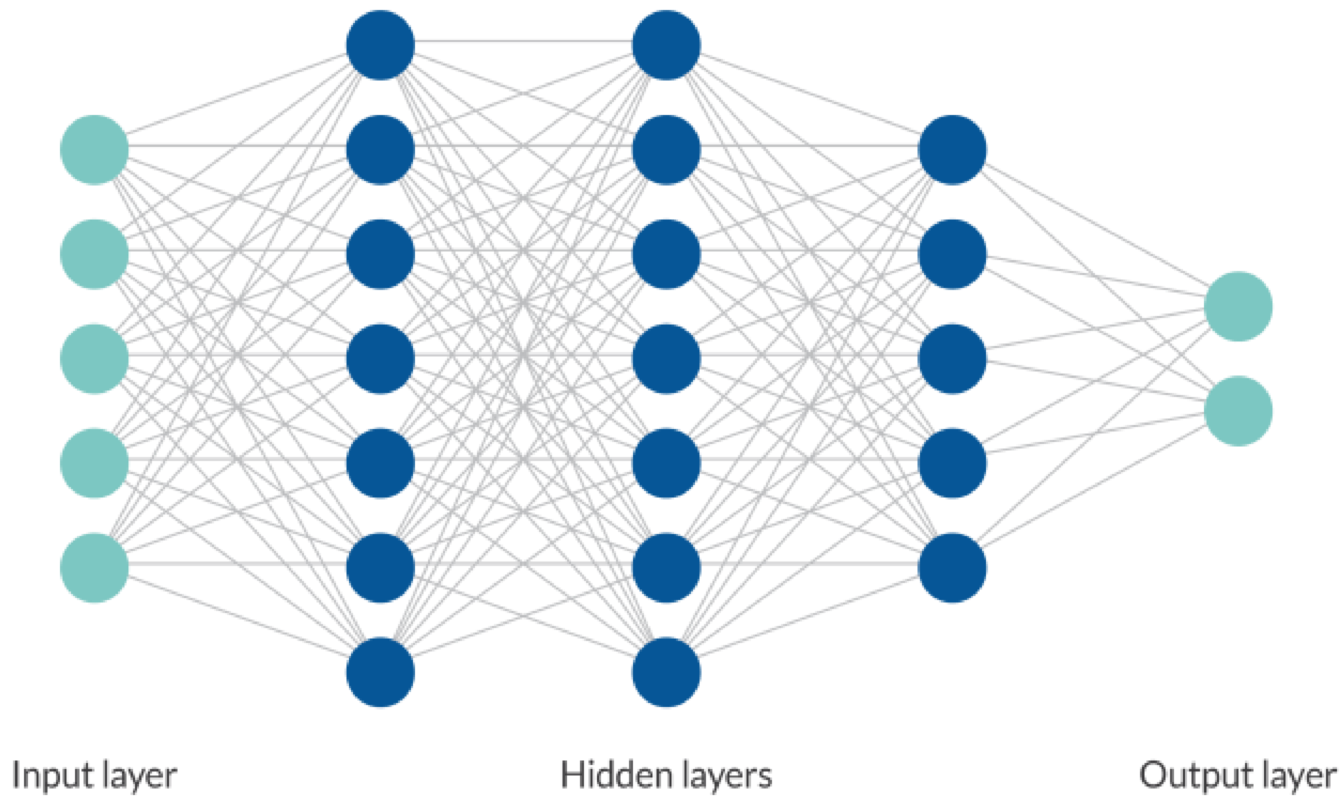
3 weeks: Math as an NLP problem (LLMs etc.)

3 weeks: Reinforcement Learning (RL) in Math

1 week: Advanced AI topics

1 week: Project Presentations

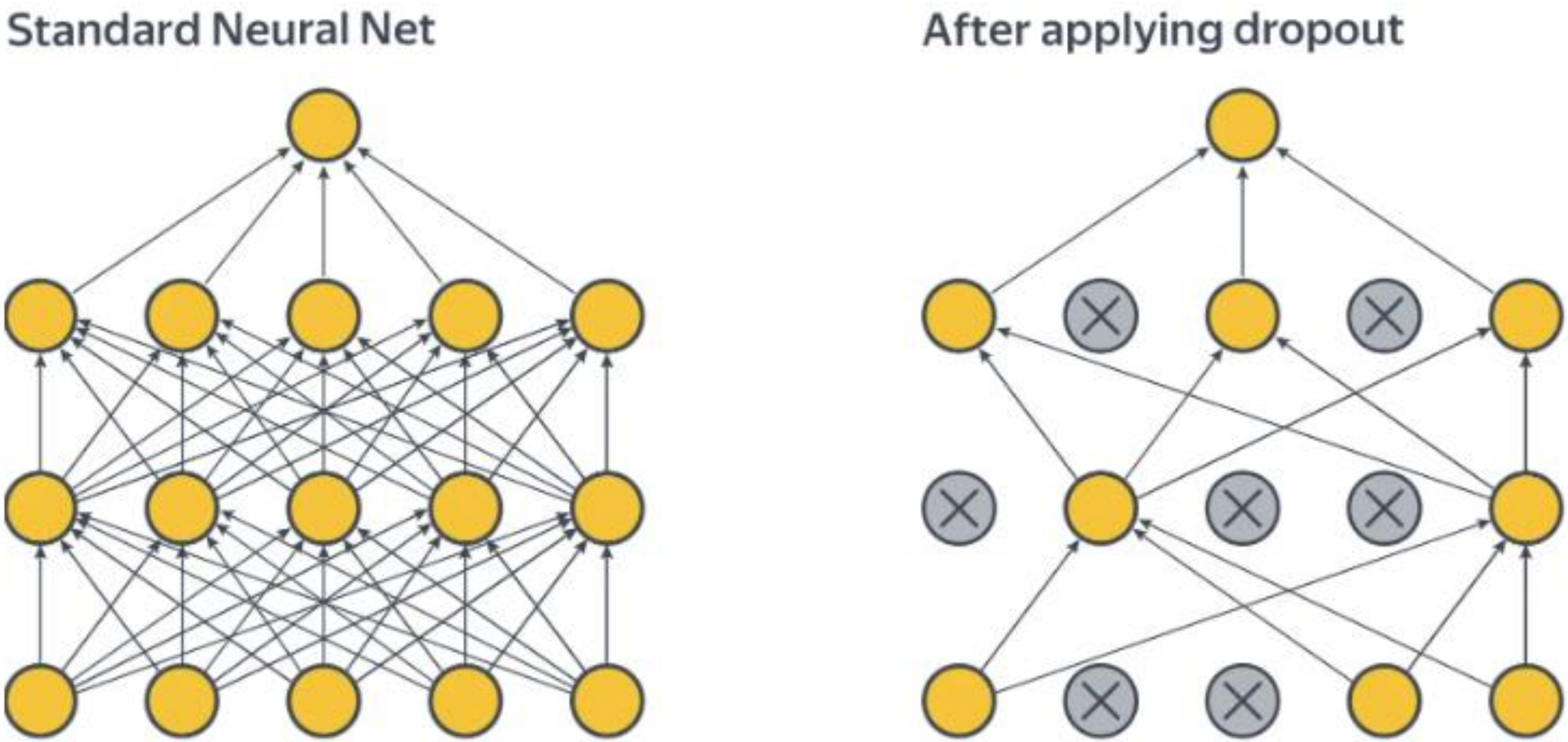
Neural Network



$$f(x) = f_L \circ \cdots \circ f_1(x), \qquad f_i(x) = \sigma_i(xW_i^T + b_i)$$

$$W_i \in \mathbb{R}^{n_{i-1} \times n_i}, \qquad b_i \in \mathbb{R}^{n_i}, \qquad x \in \mathbb{R}^{n_0}$$

Dropout



BatchNorm

There's **no theoretical guarantee**, but it has **empirically shown strong performance** across many models. Helps stabilize training and improve convergence speed.

🔄 Training Phase

1. For each batch: $X'_k = \frac{X_k - \mu}{\sqrt{\sigma^2 + \varepsilon}}$
where \mu and \sigma^2 are the **batch mean and variance**.
2. Running estimates are updated:
$\mu_* = \lambda \mu_* + (1 \ - \lambda)\mu$ and $\sigma_* = \lambda \sigma_* + (1 \ - \lambda)\sigma$
3. Then apply **learnable transformation**:

$$X_{k+1} = \gamma X'_k + \beta$$

where β and γ are **trainable parameters** (not fixed hyperparameters).

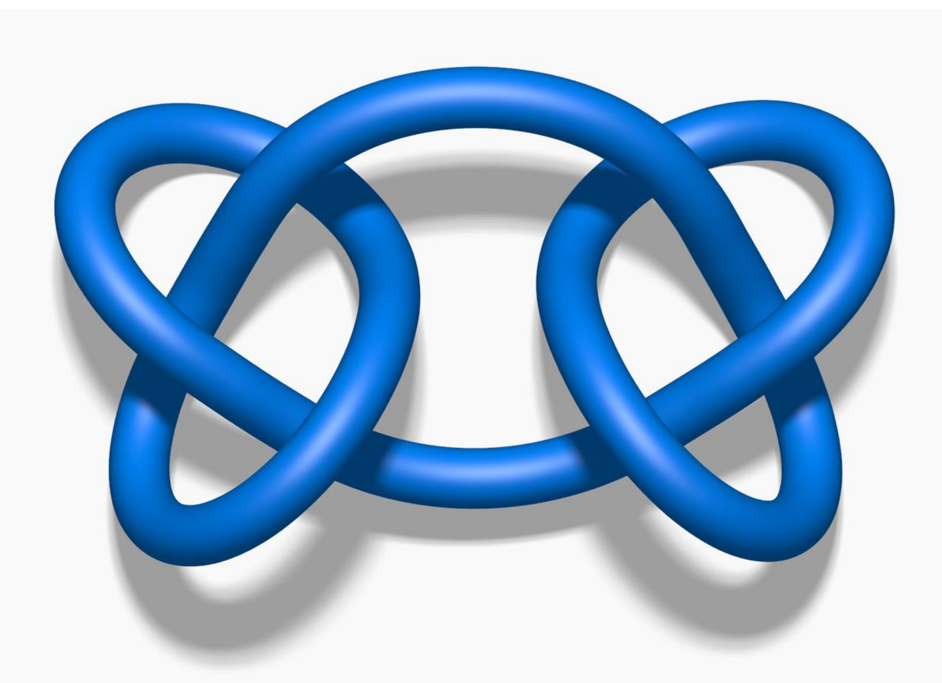
🧪 Test Phase

Use the **running estimates** μ_* , σ_*^2, *and learned* β, γ *to normalize:*

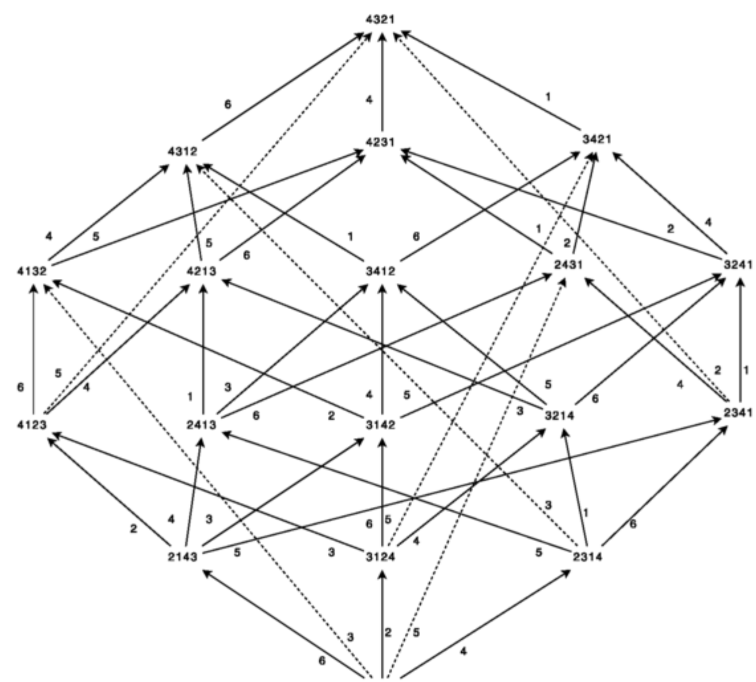
$$X'_k = \frac{X_k - \mu_*}{\sqrt{\sigma_*^2 + \varepsilon}} \rightarrow X_{k+1} = \gamma X'_k + \beta$$

Advancing mathematics by guiding human intuition with AI (2021)

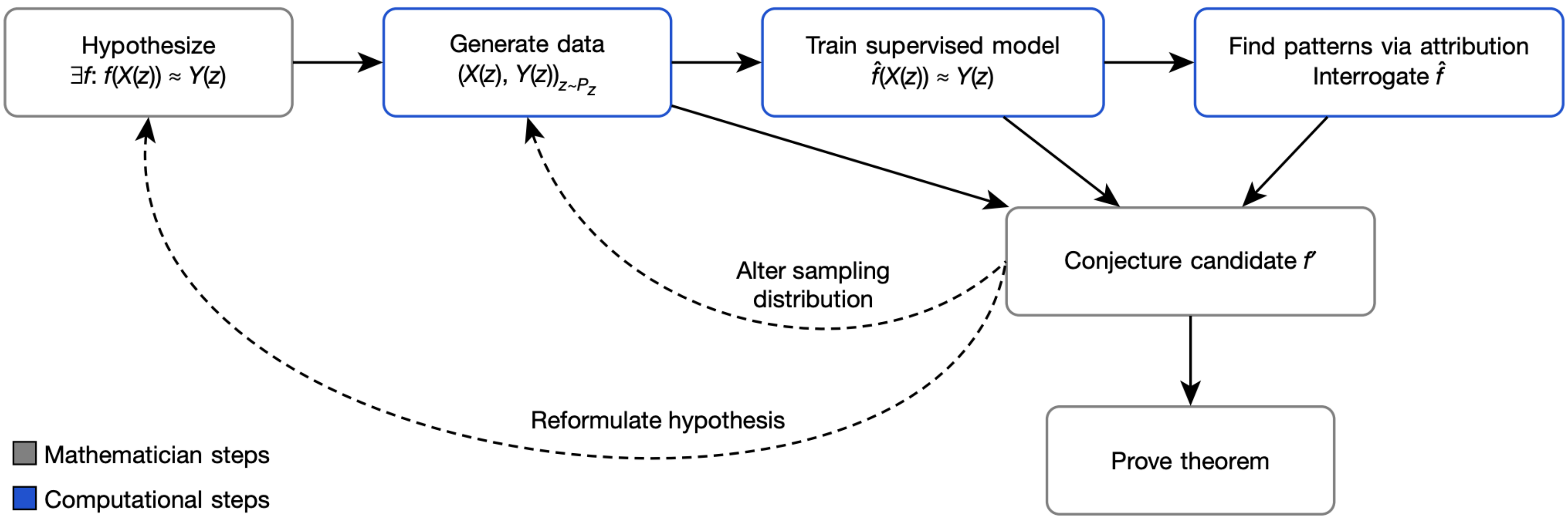
Knot Theory



Representation theory



General framework suggested by the authors



Representation theory

S_n – the group of all permutations of n elements.
Each element $\sigma \in S_n$ is a **bijection** from the set $\{1, 2, \ldots, n\}$ to itself.
For example: $\sigma = \binom{12345}{23514} \in S_5$.

This means:
$\sigma(1) = 2, \sigma(2) = 3, \sigma(3) = 5, \sigma(4) = 1, \sigma(5) = 4$

We can also represent σ by just a bottom line:

$$\sigma = \binom{12345}{23514} = (23514)$$

Representation theory

Inversions:

•The **number of inversions** in σ, denoted $i(\sigma)$, is defined by:

$i(\sigma) = \#\{k < l \mid \sigma(k) > \sigma(l)\} -$ number of **inversions**.

This measures how much the permutation “disorders” the natural order.

$$i < j$$

$$\begin{pmatrix} 1 & 2 & 3 & 4 & 5 \\ 3 & 5 & 1 & 2 & 4 \end{pmatrix}$$

$$\pi(i) > \pi(j)$$

Bruhat Order Motivation

Let $\mathrm{A}(a_{i,j}) = (I + a_{12}E_{12})\ldots(I + a_{n-1\,n}E_{n-1\,n})$ denote an upper triangular matrix with diagonal entries equal to 1, where E_{ij} is the elementary matrix with 1 in position (i,j) and $a_{ij} \in \mathrm{K}$ for $i <$ j.

$$\Omega_w = \{T \circ A(a_{i,j}) \circ \boldsymbol{w} \circ A(b_{i,j}) |, T - diagonal, a_{ij}, b_{ij} \in K\}$$

Where $b_{ij} = 0$, if the matrix $w(I + E_{i,j})w^{-1}$ is lower triangular

$$v \leq w \Leftrightarrow \Omega_v \subseteq \overline{\Omega_w}.$$

Example in GL_2

$$\Omega_e = \{\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}, b \in K, a, c \in K^*\}$$

$$\Omega_{(1,2)} = \left\{\begin{pmatrix} a & 0 \\ 0 & d \end{pmatrix}\begin{pmatrix} 1 & b \\ 0 & 1 \end{pmatrix}\begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}\begin{pmatrix} 1 & c \\ 0 & 1 \end{pmatrix}, b, c \in K, a, d \in K^*\right\} =$$
$$= \left\{\begin{pmatrix} ab & a + abc \\ d & cd \end{pmatrix}, b, c \in K, a, d \in K^*\right\}$$

As $d \rightarrow\ 0$, we can choose a, b, c such that $\Omega_e \subseteq \overline{\Omega_{(1,2)}}$.

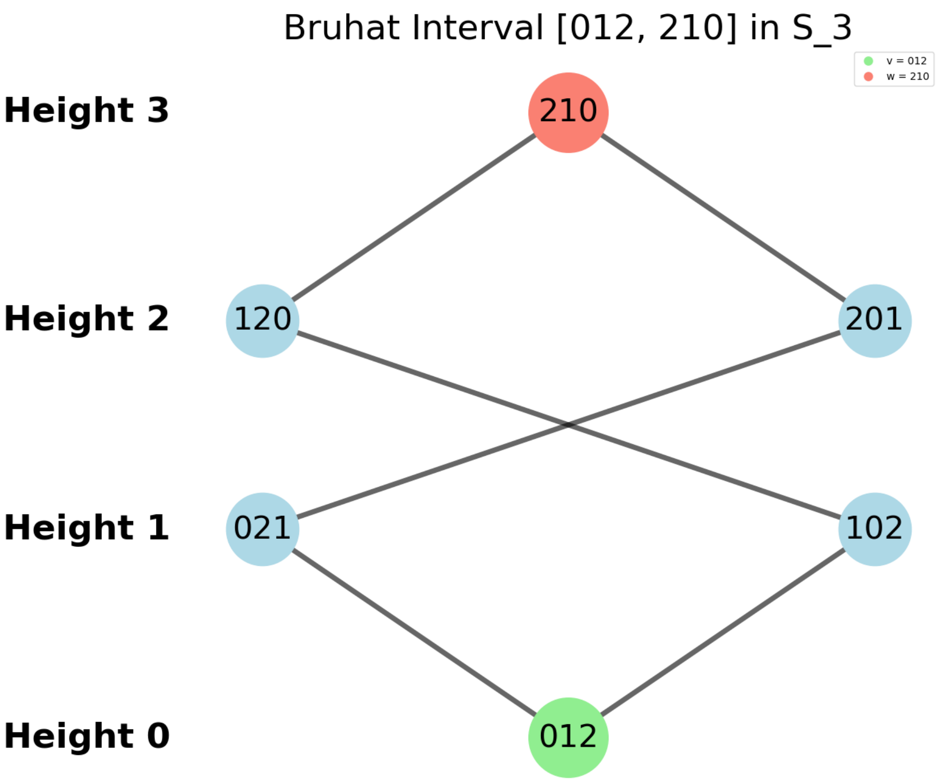
Representation theory

The **Bruhat graph** is a **directed graph** where:
Vertices: $V = S_n$
Edges:

$$\sigma \leftarrow \tau \quad \Longleftrightarrow$$

$\tau = \sigma \circ t, and \ i(\sigma) < i(\tau)$.
Where t is a simple transposition $(s, s+1)$.

Bruhat Order:
The **Bruhat order** $\leq$ is the **transitive closure** of the relation $\leftarrow$ in the Bruhat graph.



Representation theory

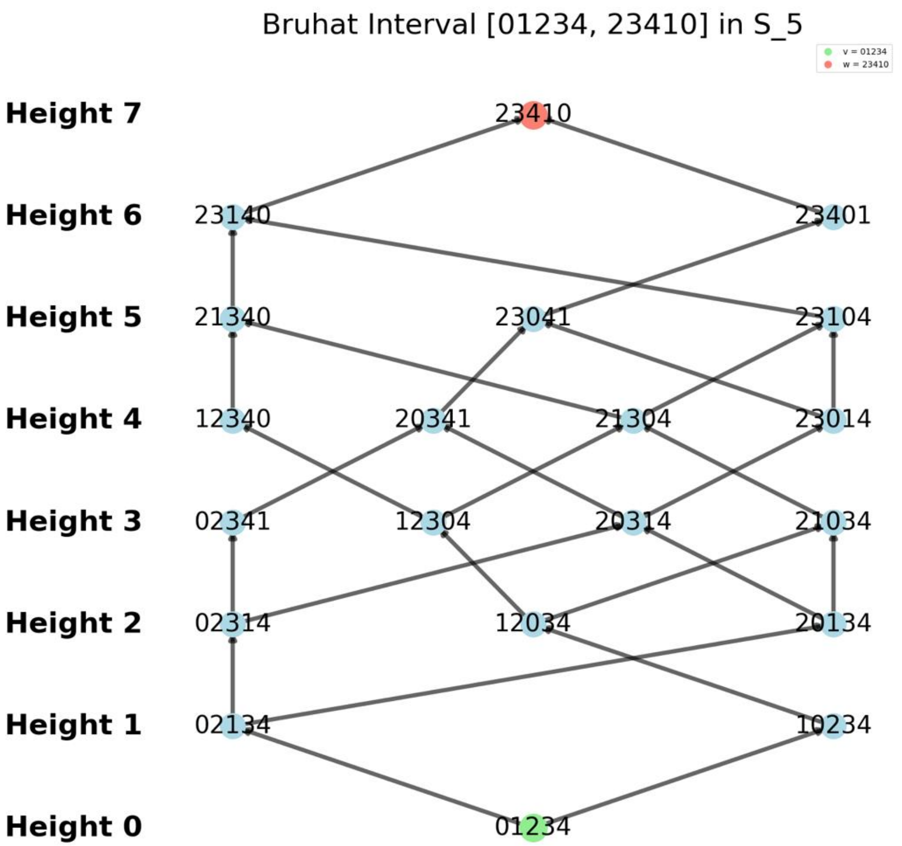
A **Bruhat interval** is the **subgraph** of the Bruhat graph **induced by a pair** $\sigma \leq \tau$.

- It includes **all permutations** $\chi \in S_n$ such that:

$$[\sigma, \tau] = \{ \chi \in S_n \mid \sigma \leq \chi \leq \tau \}$$

This interval forms a **subgraph** within the full Bruhat order on S_n

The interval [01234, 23410]:



KL polynomials

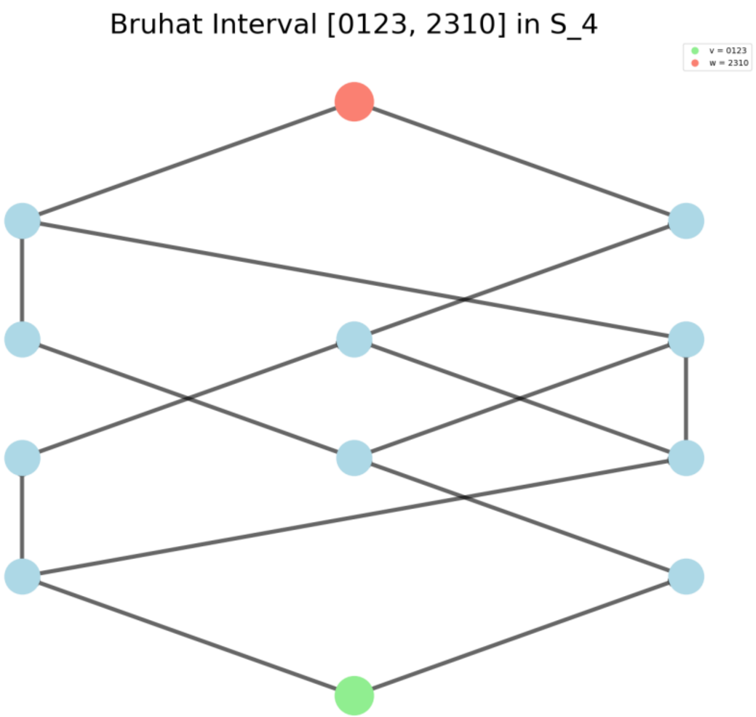
For any pair of elements σ, τ, Kazhdan and Lusztig defined a polynomial:

$$P_{\sigma,\tau}(q) \in \mathbb{Z}[q]$$

Defined **inductively**, starting from the identity element. At each step, the computation may depend on previously calculated polynomials. Arise from **Hecke algebra theory**.

Combinatorial Invariance Conjecture

Conjecture:
The KL polynomial $P_{\sigma,\tau}(q)$ depends only on the structure of Bruhat graph of the interval $[\sigma, \tau]$.

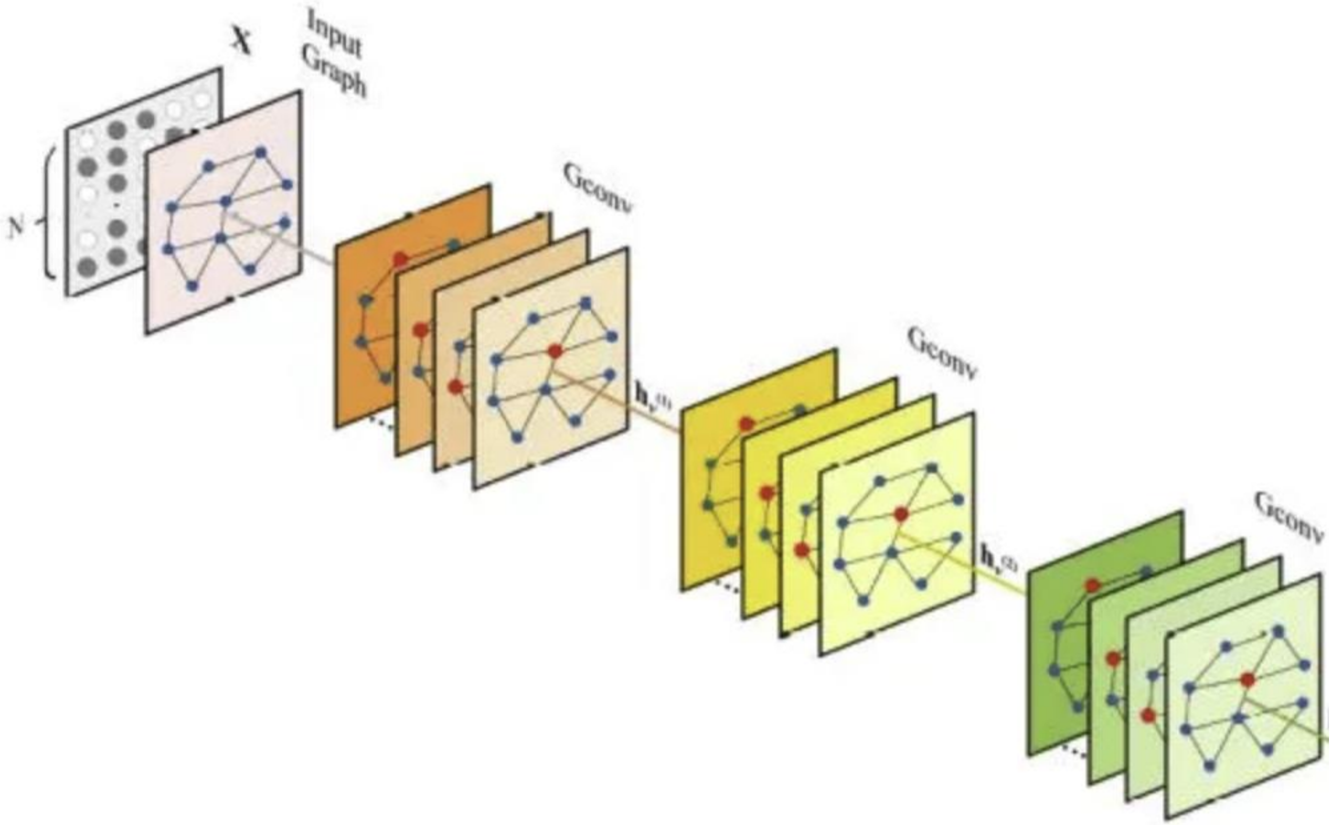


$$\longrightarrow \quad P_{\sigma,\tau}(q)$$

Combinatorial Invariance Conjecture

z: **Pair of permutations**	$X(z)$: **Unlabelled Bruhat interval**	$Y(z)$: **KL polynomial**
(03214), (34201)		$1 + q^2$
(021435), (240513)		$1 + 2q + q^2$

Graph Neural Network

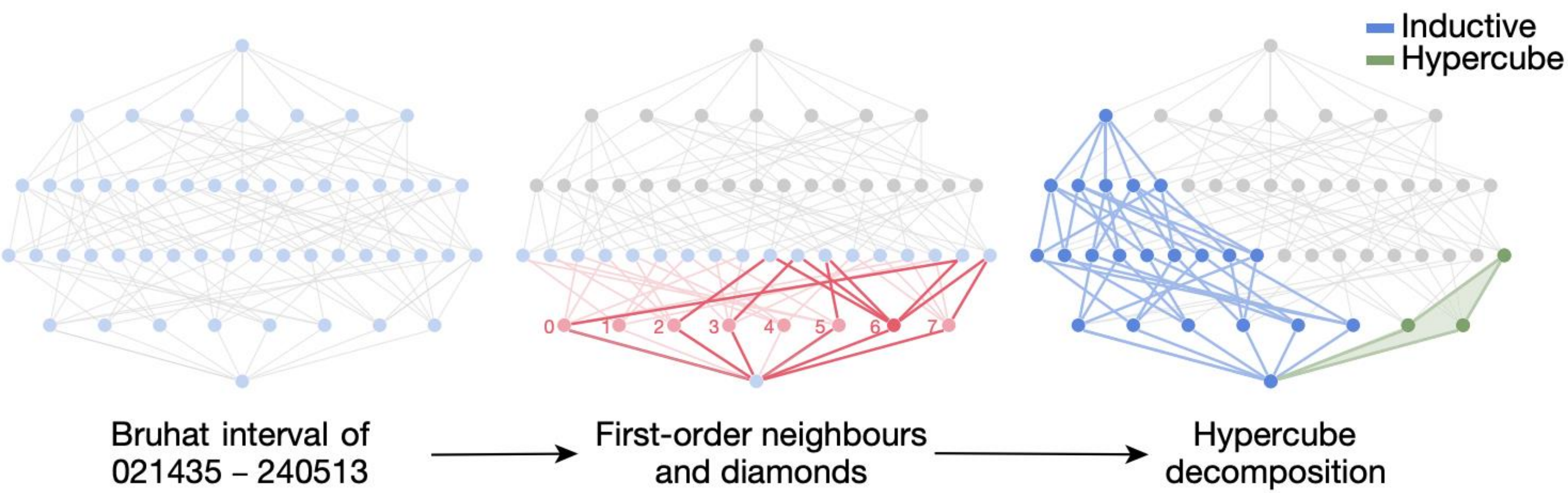


Combinatorial Invariance Conjecture

•A **neural network** was trained to compute Kazhdan–Lusztig polynomials for labeled Bruhat intervals.
•The network achieved **high accuracy**, uncovering interesting structure in the input data.

Main Theorem (Conjectured & Proved):

Every labeled Bruhat interval admits a **canonical hypercube decomposition** along its **extremal reflections?**, from which the **KL polynomial** can be **directly computed**.



Combinatorial Invariance Conjecture

The neural model suggested something even deeper:
It seems that the KL polynomial can be reconstructed
from **any hypercube decomposition**, not just the canonical one.
This observation was **experimentally verified** for S_n with $n < 8$.

Second Conjecture (Open):

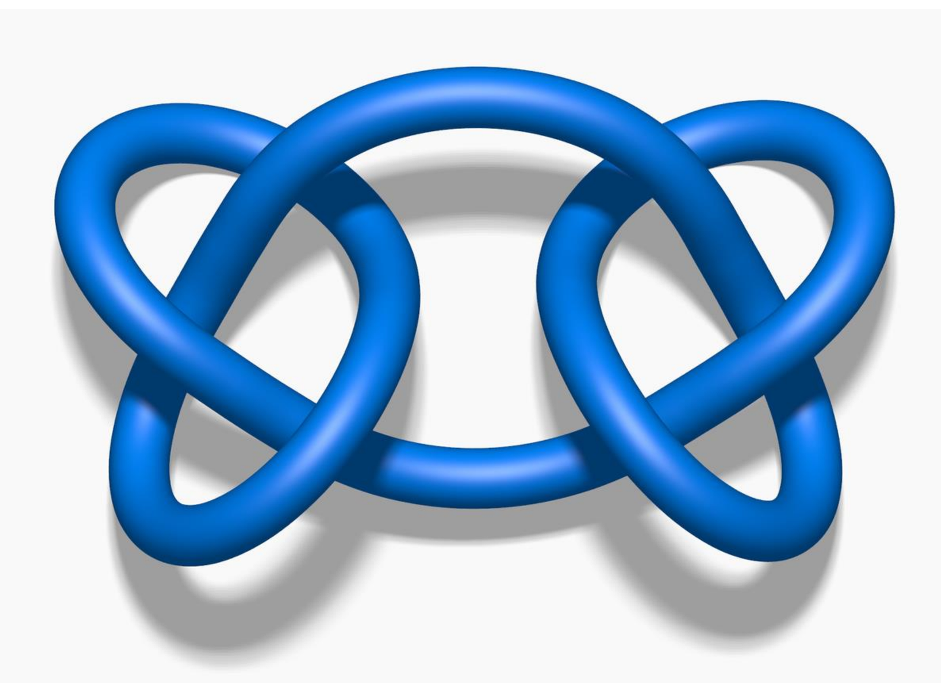
The **Kazhdan–Lusztig polynomial** of an **unlabeled Bruhat interval**
can be computed from **any hypercube decomposition**
using the same formula as in the labeled case.

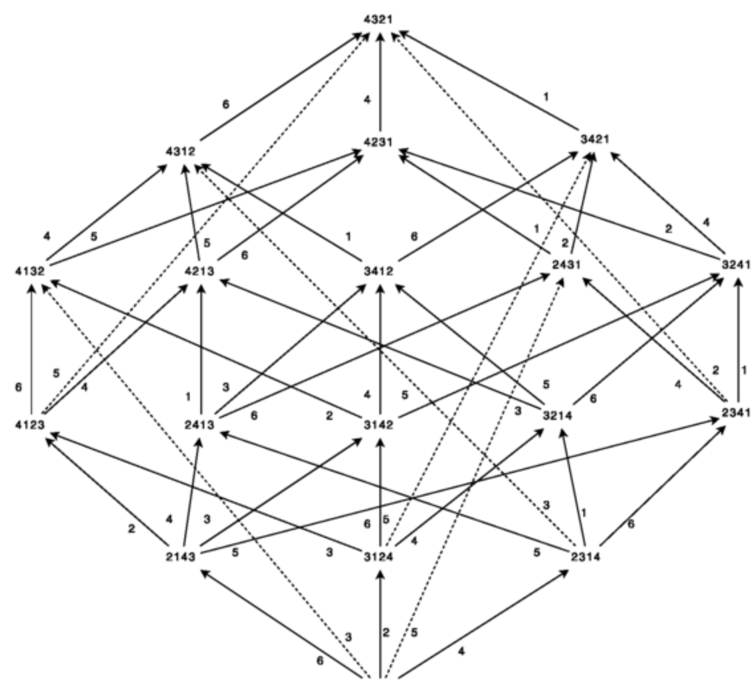
✅ **If proved**, this would **resolve the Combinatorial Invariance Conjecture**

Advancing mathematics by guiding human intuition with AI (2021)

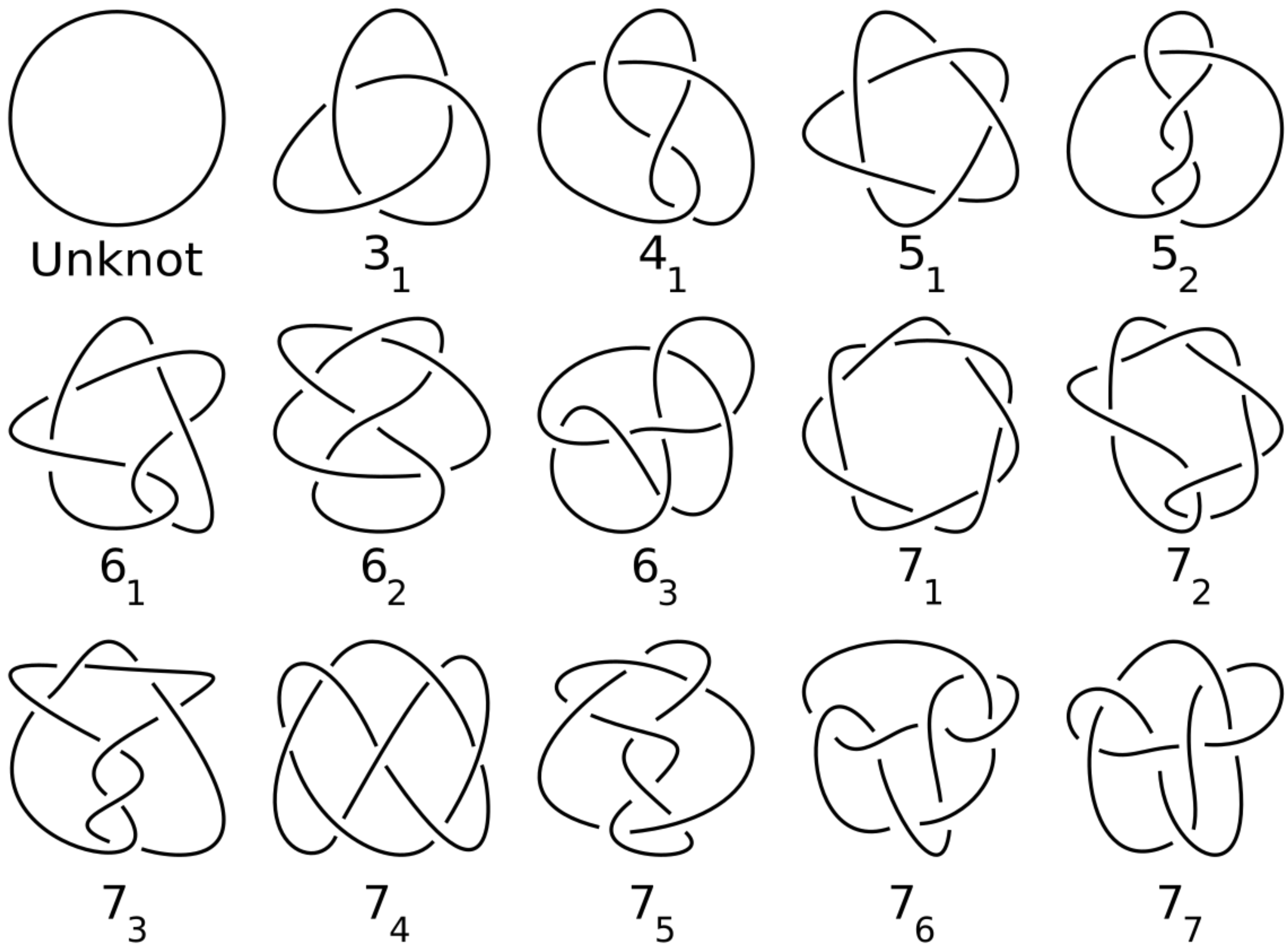
Knot theory

~~Representation theory~~





Knot Theory

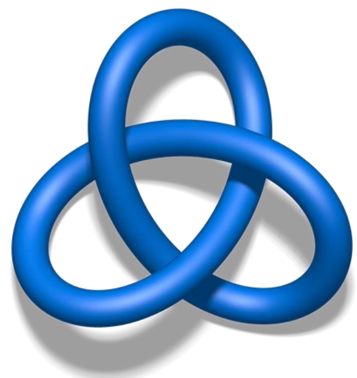


Knot Theory

🧵 What Is a Knot (Formally)?

A **knot** is a smooth embedding of the circle:

$$S^1 \to \mathbb{R}^3$$



Knots have a lot of different invariants of different nature and it is important to find relations between them.

Knot Theory

🧬 **What is the Signature of a Knot?**

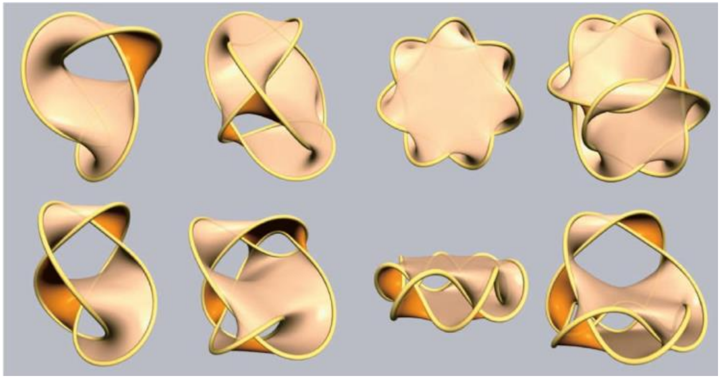
•We can construct a **Seifert surface** — a smooth, oriented surface whose boundary is the knot.

•From this surface, we build a matrix called the **Seifert matrix**, using information from the surface's homology.

•The **signature of the knot** is defined as the **signature of this matrix** (i.e., the number of positive eigenvalues minus the number of negative ones).

✅ **Key Point**

The **signature** is a **knot invariant** — it stays the same no matter which Seifert surface is used.



$$A = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -1 \end{bmatrix}$$

Knot Theory

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	$0.7381 + 0.8831i$	...	–2	$t - t^2 + 2t^3 - t^4 + t^5 - t^6$	...
	3.1640	0.1560	$-0.7237 + 1.0160i$	...	0	$t^{-2} - t^{-1} + 2 - 2t + t^2 - t^3 + t^4$	...

Fig. 2 | Examples of invariants for three hyperbolic knots. We hypothesized that there was a previously undiscovered relationship between the geometric and algebraic invariants.

🎯 Goal of the Study

The authors aim to explore whether there exists a hidden relationship between **algebraic** and **geometric** knot invariants.

Specifically, they focus on predicting the **signature** of a knot — an algebraic invariant — using **geometric features** of the knot.

We'll see how they approached this problem in today's notebook!

Geometric Invariants

Let $K \subset S^3$ be a hyperbolic knot. Then its complement $M = S^3 \setminus K$ has a complete finite-volume hyperbolic metric. From this structure, we get canonical geometric invariants:

- **Hyperbolic Volume:**

Volume of M with the hyperbolic metric. Topological invariant.

- **Injectivity Radius**:

Half the length of the shortest nontrivial loop through point p \in M. Measures local geometric "thickness."

- **Meridional and Longitudinal Translations**:

Complex translation vectors describing how the outer torus wraps around the knot.

Slope of a knot is defined by authors as a real part of ratio of **Meridional and Longitudinal Translation.**

Knot Theory

By applying machine learning techniques, the authors initially formulated an incorrect mathematical hypothesis:

$$|2\sigma(K) - \mathrm{slope}(K)| < c_1 \mathrm{vol}(K) + c_2$$

by analyzing the behavior of counterexamples, they were able to refine their understanding — ultimately leading to an updated hypothesis, which they then **proved:**

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