Using neural networks to search for interesting mathematical objects

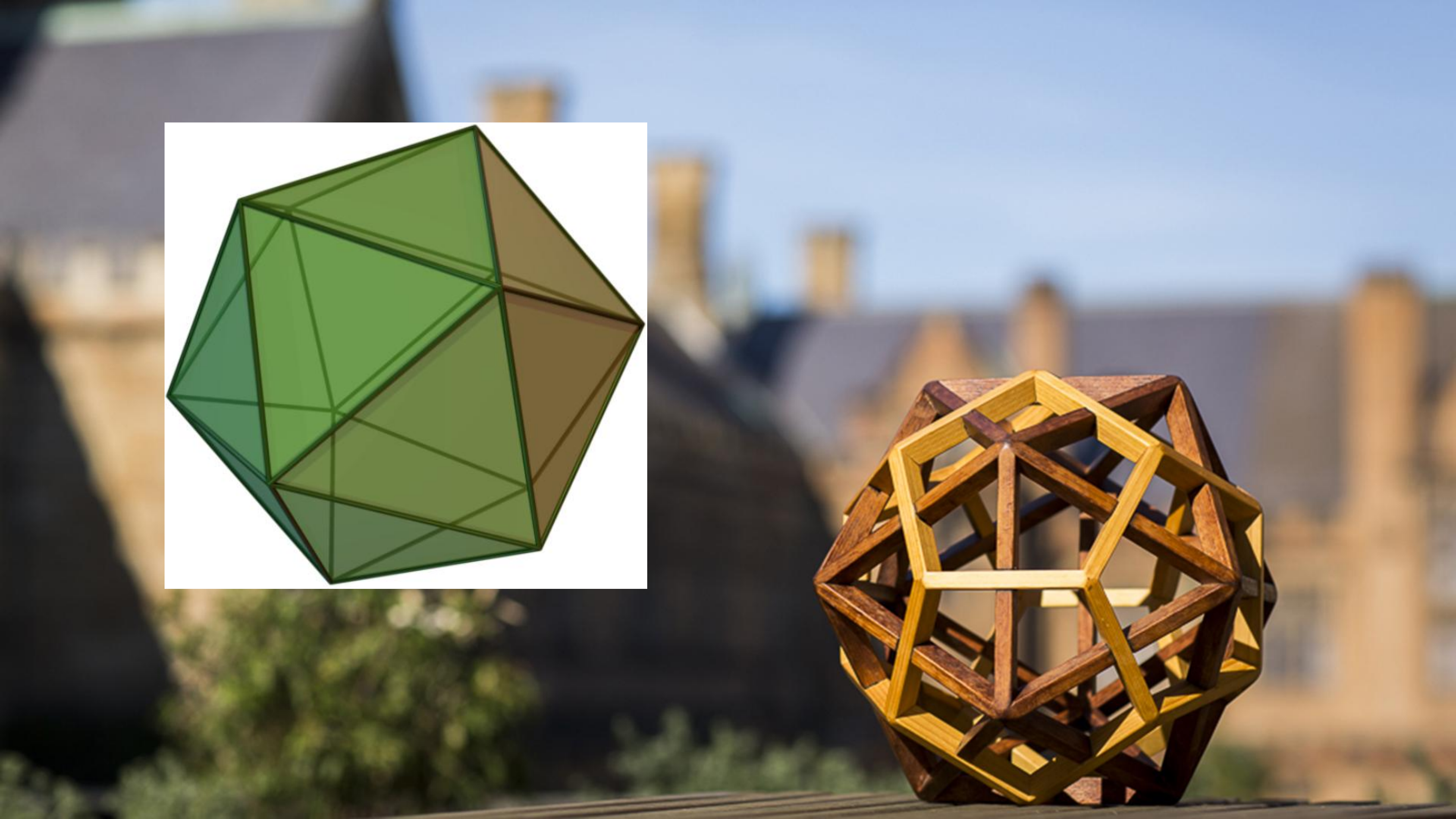
Machine Learning and Mathematics, KIAS

Geordie Williamson

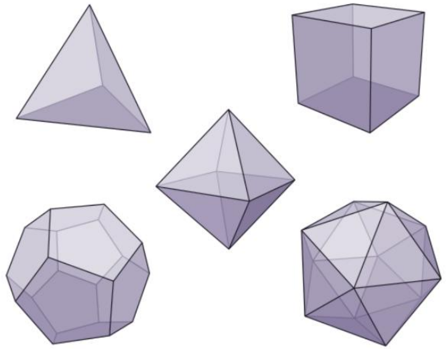
THE UNIVERSITY OF SYDNEY

Mathematical Research Institute

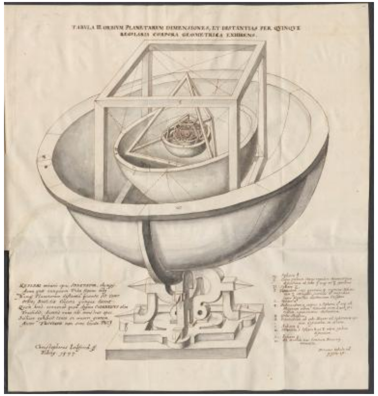
CRICOS 00026A TEQSA PRV12057



The icosahedron throughout the ages

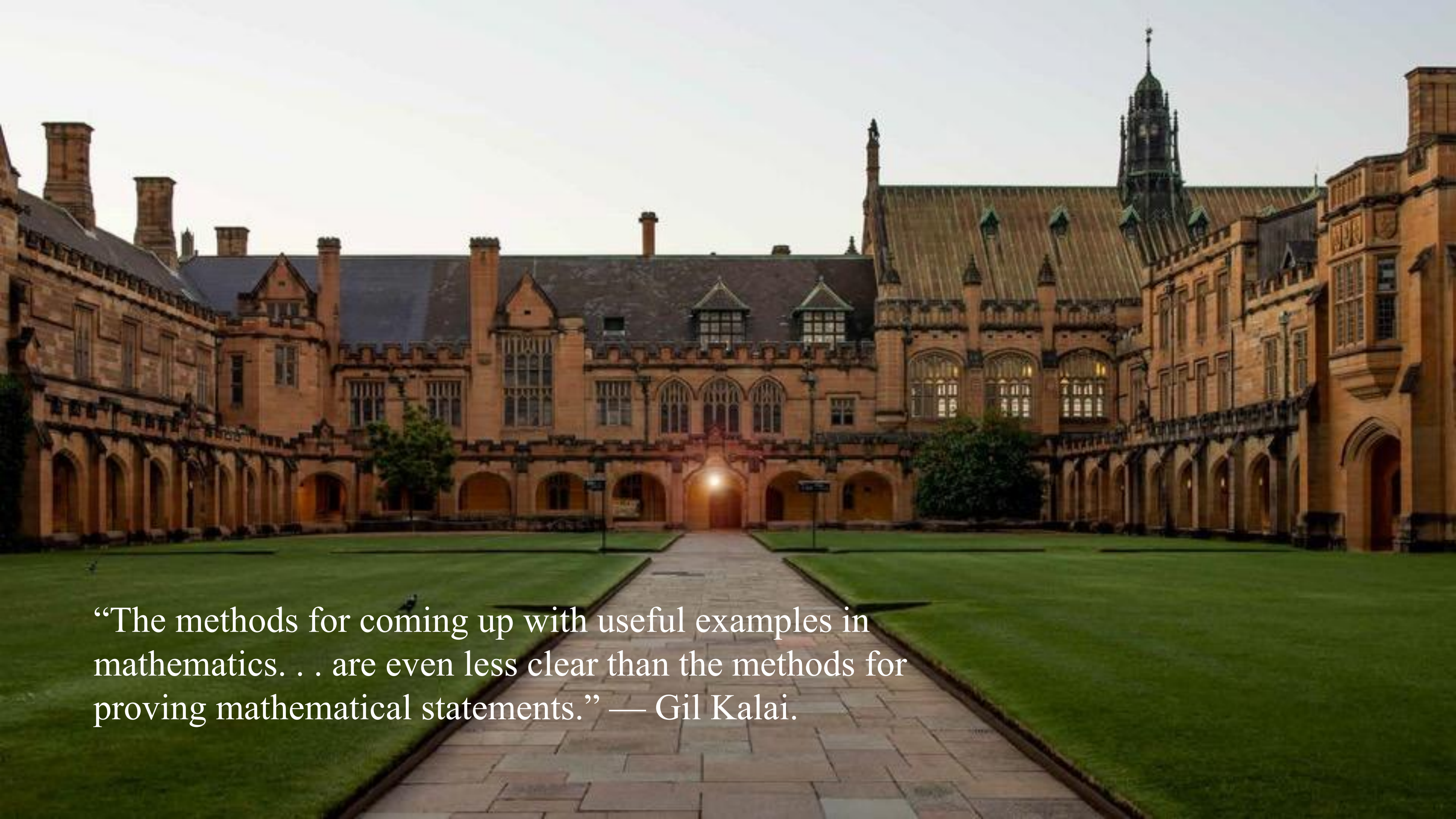


Plato's *Timaeus* (c. 360 B.C)
Euclid's *Elements* (c. 300 B.C)



Kepler's
Mysterium Cosmographicum
(1596).

Poincaré's homology sphere, which led to the Poincaré conjecture (1904), solved by Perelman (2002).



“The methods for coming up with useful examples in mathematics. . . are even less clear than the methods for proving mathematical statements.” — Gil Kalai.

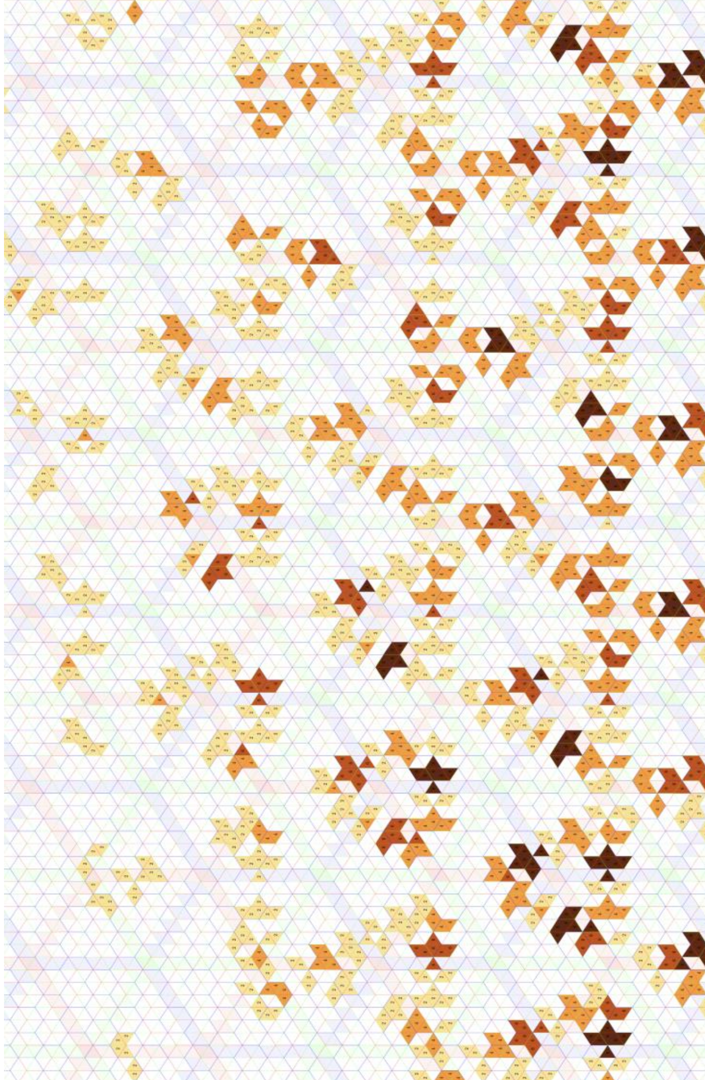
On the search for examples

Alongside the search for a proof, the search for interesting examples is an important part of the mathematical process.

Sometimes one example can give rise to an entire field (e.g. the Mandelbrot set, the KdV equation,...).

The absence of general methods to find examples is often a blind spot in many current “AI for math” initiatives.

Interesting examples are often surprising and hard to find. Neural networks detect general statistical patterns, so it is a priori surprising that they can be helpful.



Plan of talk

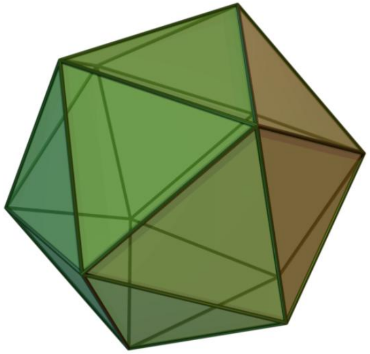
- A story about simple groups
- Constructions in combinatorics using transformers
- Finding interesting polytopes using transformers
- Attempting to attack the Jones unknot problem with a neural network.

Finite simple groups and computers

A historical example of human computer interaction

Finite simple groups

The building blocks of finite symmetry.
"Simple groups are to (symmetry) groups, as primes are to whole numbers."



Example: There are 60 rotational symmetries of the icosahedron, and these provide a simple group.

Finite simple groups

The building blocks of finite symmetry.

1832-1962: Galois, Jordan, Brauer, Chevalley, Steinberg, Thompson, ...
"most finite simple groups are close relatives of continuous symmetry".

Algebraists in the audience will know that we have names for them:
$\mathbb{Z}/p\mathbb{Z}$, A_n, $PGL_n(\mathbb{F}_q)$, …, $\mathrm{Sz}(2^{2n+1})$, … , $E_8(\mathbb{F}_q)$.

1861, 1873: Mathieu discovered 5 "sporadic" groups of sizes
7920, 95040, 443520, 10200960 and 244823040.

Are there any others? Until 1963 this was the full list.

Zvonimir Janko

- Moved to Canberra in the early 1960s as a research fellow.

- Whilst attempting to reconstruct an argument by Thompson that a certain configuration could not occur, he became convinced that it *could*.

$$A = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}, \qquad B = \begin{pmatrix} -3 & 2 & -1 & -1 & -3 & -1 & -3 \\ -2 & 1 & 1 & 3 & 1 & 3 & 3 \\ -1 & -1 & -3 & -1 & -3 & -3 & 2 \\ -1 & -3 & -1 & -3 & -3 & 2 & -1 \\ -3 & -1 & -3 & -3 & 2 & -1 & -1 \\ 1 & 3 & 3 & -2 & 1 & 1 & 3 \\ 3 & 3 & -2 & 1 & 1 & 3 & 1 \end{pmatrix}.$$

(7×7 matrices with coefficients in $\mathbb{Z}/11\mathbb{Z}$))

Zvonimir Janko

- Moved to Canberra in the early 1960s as a research fellow.

- Whilst attempting to reconstruct an argument by Thompson that a certain configuration could not occur, he became convinced that it *could*.

- Janko assigned two students (Terry Gagen and Martin Ward) the task of verifying that two matrices generated a group of order 175560. (They worked on an IBM 1620, with 40K of memory.)

> By the end of the night—Terry's memory is that dawn was just breaking—they had found elements of orders 3, 5, 7, 11, 19 and with the other information they had, they knew that the new simple group existed as Janko had suspected.
>
> -- D. Taylor and T. Gagen

Zvonimir Janko

- Moved to Canberra in the early 1960s as a research fellow.

- Whilst attempting to reconstruct an argument by Thompson that a certain configuration could not occur, he became convinced that it *could*.

- Janko assigned two students (Terry Gagen and Martin Ward) the task of verifying that two matrices generated a group of order 175560. (They worked on an IBM 1620, with 40K of memory.)

- The first new sporadic simple group in over a century!

The monster simple group

Over the next decade a further 18 sporadic groups were found.

The largest is the monster simple group. Its order is

808017424794512875886459904961710757005754368000000000

which is roughly the number of elementary particles in Jupiter.

It was conjectured to exist in 1975 by Fischer, and proved to exist by Griess in 1982. Computation in the monster is a major challenge.

Martin Seysen's *mmgroup* package (2022) is a major advance.

The classification of finite simple groups

It is now generally agreed that all finite simple groups are known. They consist of the known families, together with the 26 sporadic groups, including the Mathieu, Janko and monster simple groups.

The road to their discovery combined human ingenuity with computation.

Would we have the classification today without computers?

The monster resisted computation for nearly forty years.

For one sporadic simple group (the O'Nan simple group) we have no existence proof which *doesn't* involve computers!

Zvonimir Janko

The discovery was a miracle?

I was lucky. You can discover something great only if you are critical of some works of the greatest mathematicians. If they make a mistake, a mistake that of one of the greatest mathematicians of the time, what is the probability of one in a billion, that you do not have a lot of possibilities. However, this happened.

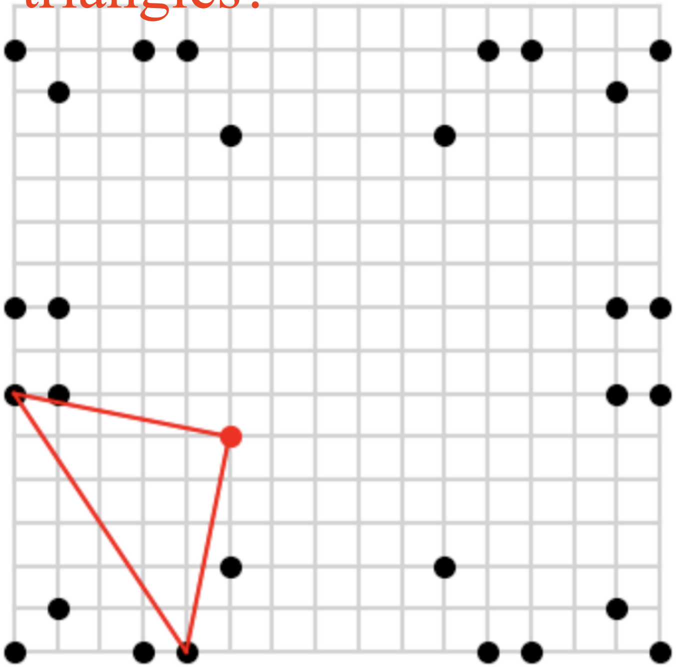
--- from an interview with Janko by Pero Zelenika.

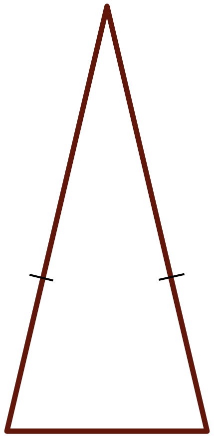
disks. The number of seconds in a year is a little more than 3×10^7. It is plausible to imagine that representatives for the orbits of $K \times K$ acting on the set of formal group elements can be found and processed at an average rate of at least 10 per second. These rough calculations indicate that the construction of the monster is technically feasible but not yet economically justifiable. I doubt that anyone would argue that the existence or nonexistence of the monster is of such importance to the mathematical community that it warrants devoting the resources of a major computer center for the better part of a year to settle the matter. However, it should be possible to refine the techniques sufficiently so that the question can be answered at a price which can be justified.

-- From Sims, *A method of constructing a group from a subgroup,* 1978.

Example 1: Extremal combinatorics

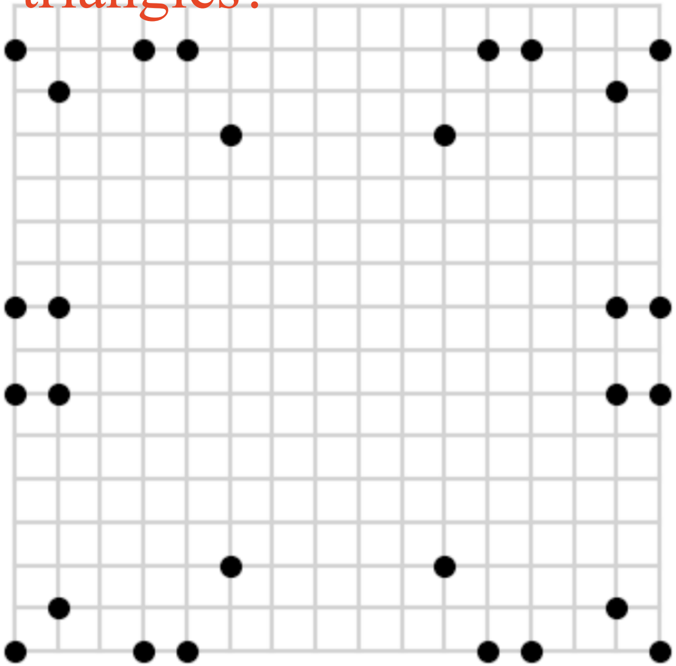
What’s the largest subset of an $n \times n$ grid without isosceles triangles?





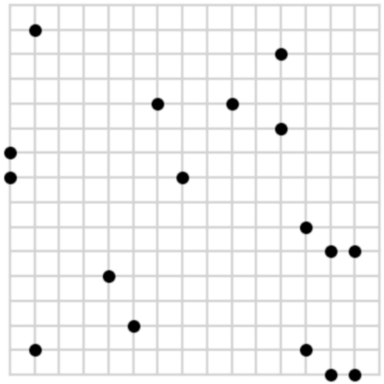
an *isosceles* triangle

What's the largest subset of an $n \times n$ grid without isosceles triangles?

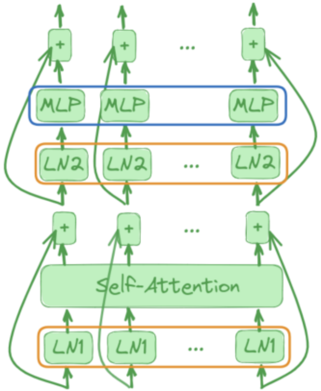


- A $2d$ variant of a classic problem in number theory and additive combinatorics ("progression free subsets", Green–Tao theorem).
- Little is known about optimal solutions.
- Hard to produce examples which are close to extremal.
- Highly addictive problem!

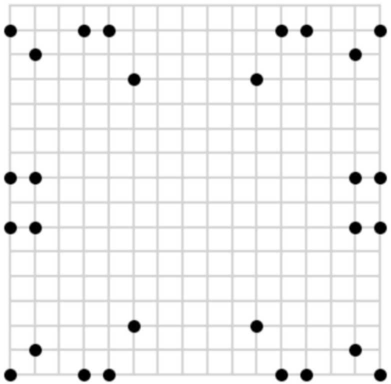
Learning from good solutions



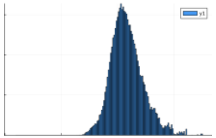
We generate many examples by local search (randomly adding points, without creating isosceles triangles).
We take **top 10%.**



Each construction gives token sequence. These are used to train a GPT-2 style transformer. We then sample from the transformer to produce new examples.



A tiny percentage of samples do better than the training set under local search. Amazingly, one often arrives at a near optimal solution.



Can Iterate!

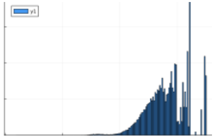
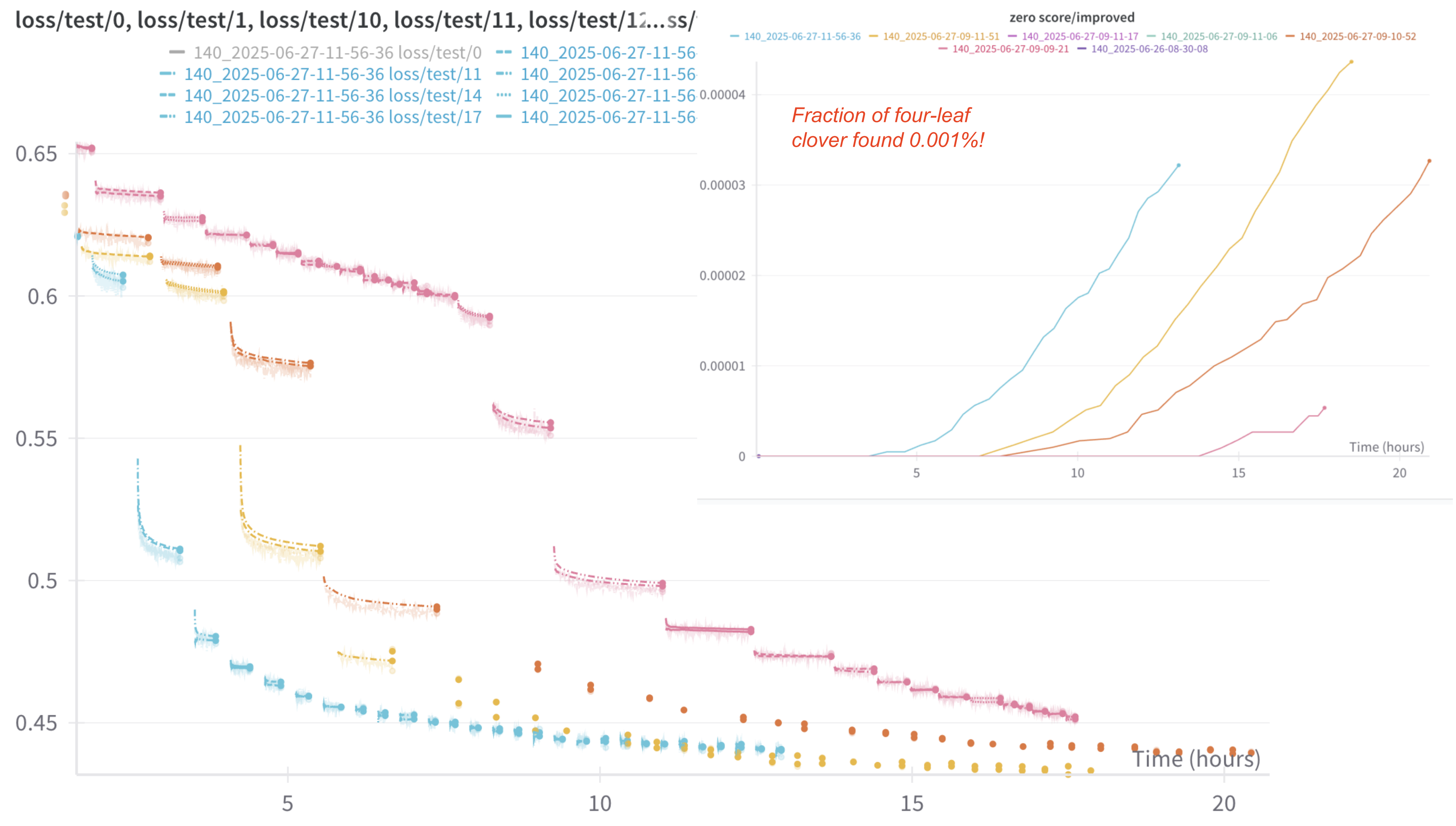


Image ©: dataflowr

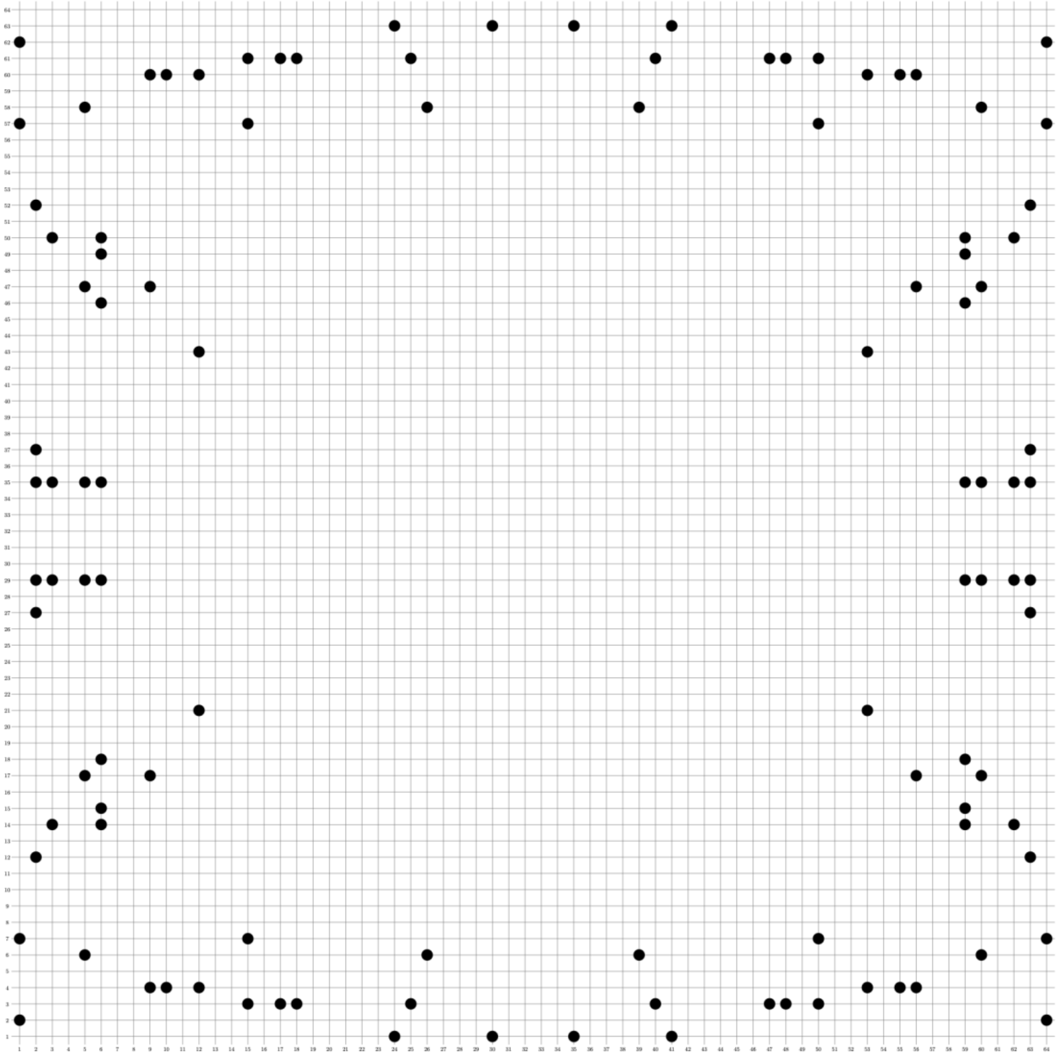

loss/test/0, loss/test/1, loss/test/10, loss/test/11, loss/test/12...ss/
140_2025-06-27-11-56-36 loss/test/0
140_2025-06-27-11-56-36 loss/test/11
140_2025-06-27-11-56-36 loss/test/14
140_2025-06-27-11-56-36 loss/test/17
zero score/improved
140_2025-06-27-11-56-36
140_2025-06-27-09-11-51
140_2025-06-27-09-11-17
140_2025-06-27-09-11-06
140_2025-06-27-09-10-52
140_2025-06-27-09-09-21
140_2025-06-26-08-30-08
Fraction of four-leaf clover found 0.001%!
Time (hours)
Time (hours)

Transformers as generators

This technique generates best known solutions to several problems in extremal combinatorics.

Recent applications of Ramos and Sun to conjectures in graph theory.

Charton, Ellenberg, Wagner, W.
PatternBoost: Constructions in Mathematics with a Little Help from AI
https://arxiv.org/abs/2411.00566

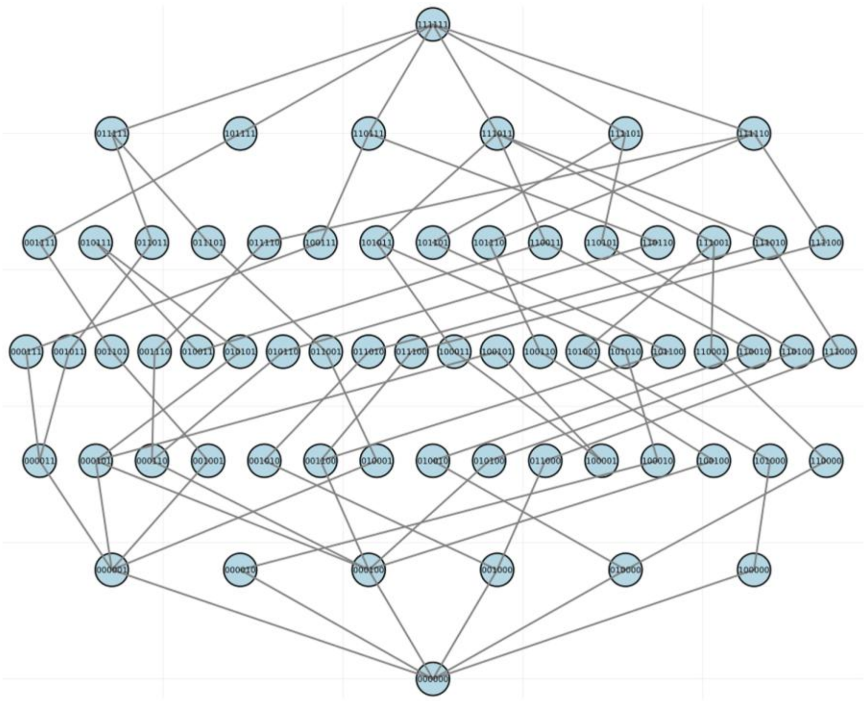


Transformers as generators

This technique works well on some problems and struggles on others. We are able ot find the the best known solutions to several problems in extremal combinatorics.

We are also able to disprove a 30-year old conjecture of Niali Graham about the graph of the hypercube:

What is the smallest subgraph of the hypercube which still has diameter n?



81 edges vs. conjectured 82!

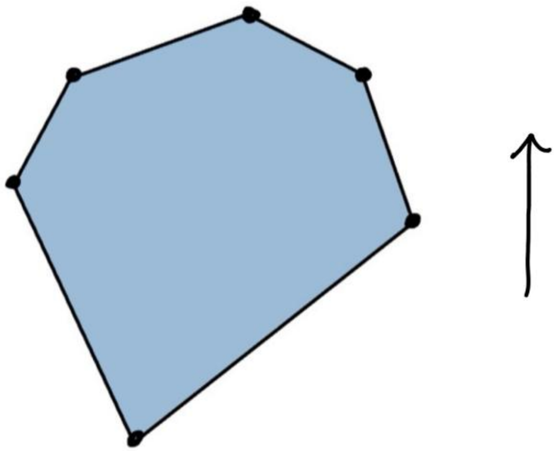
Linear programming and Hirsch conjecture

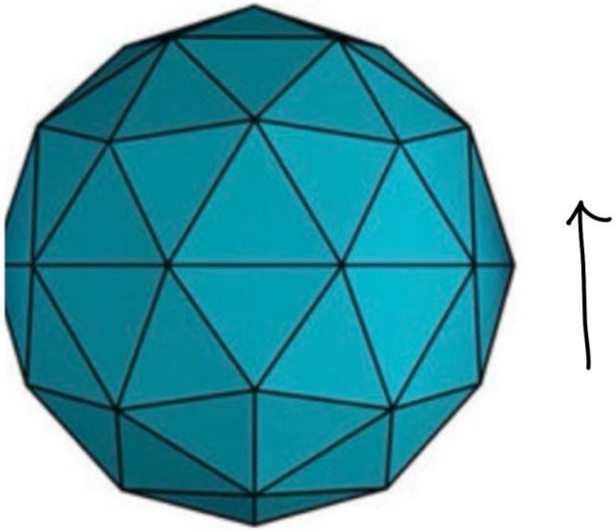
Neural networks guiding search on a hard problem

Linear programming

- A core problem in optimization.

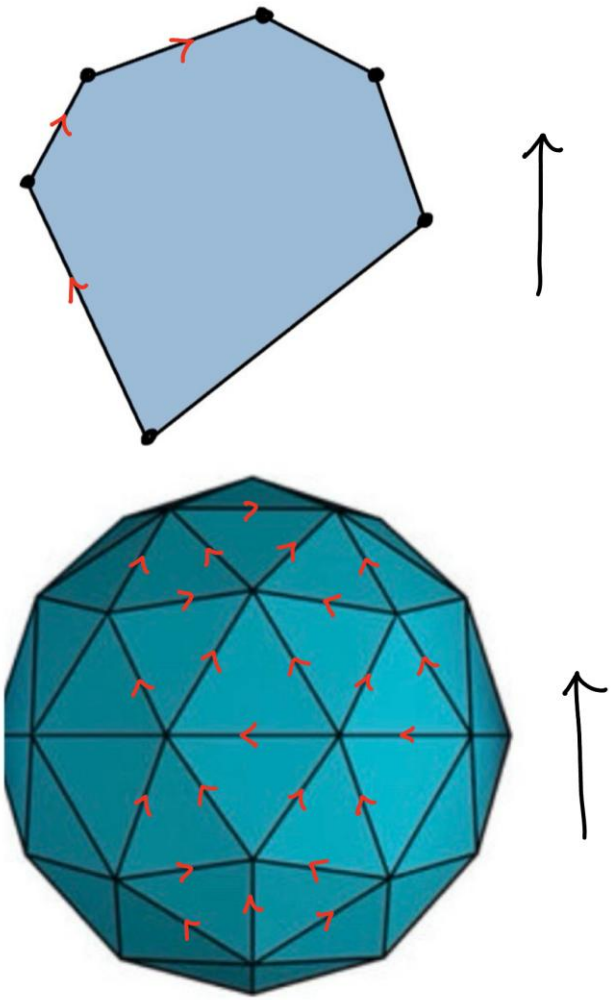
- Used > 10 million times a day.





Linear programming

- A core problem in optimization.
- Used > 10 million times a day.
- The *simplex method* is extremely efficient at solving it in practice, but we don't know why.
- *Hirsch problem*: How efficiently can we navigate the vertices of a polytope, if we are only allowed to traverse edges?

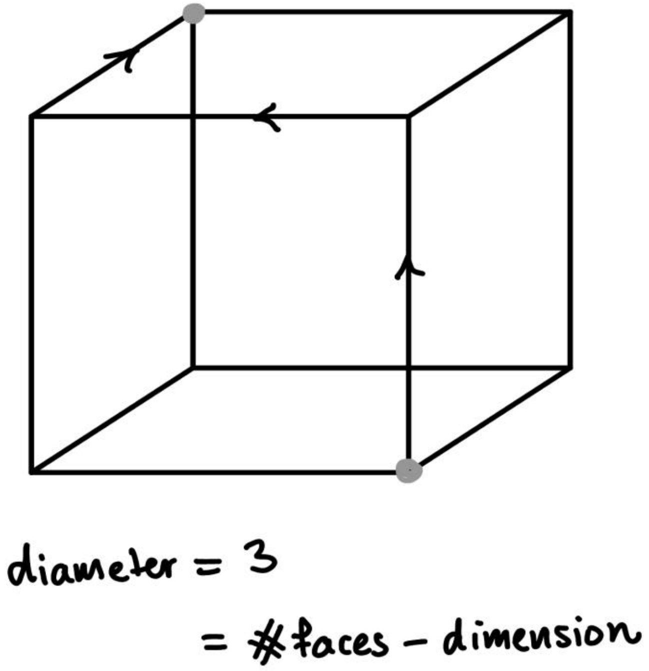


The width of polytopes

Hirsch conjecture (1957): The vertex-edge graph of any polytope with *n* facets in dimension *d* has diameter at most *n-d*.

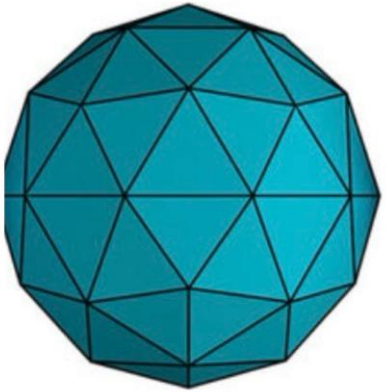
Santos (2012): There exists a 43-dimensional polytope with 86 facets of diameter at least 44 > 43 = 86 – 43.
Thus, the Hirsch conjecture is *false!*

Santos reduced this problem to an extremely difficult problem in 5 dimensions. Very few other examples known. No examples which don't use Santos' 5d trick are known.



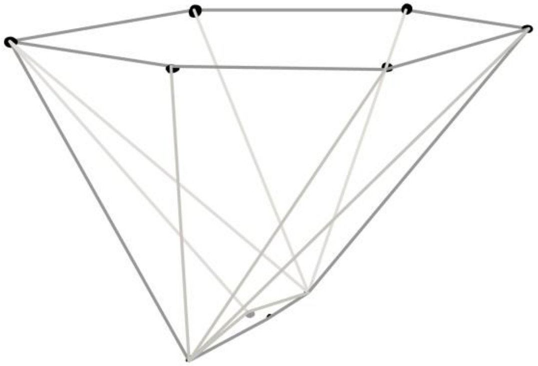
The width of polytopes

Hirsch conjecture (1957): The vertex-edge graph of any polytope with *n* facets in dimension *d* has diameter at most *n-d*.



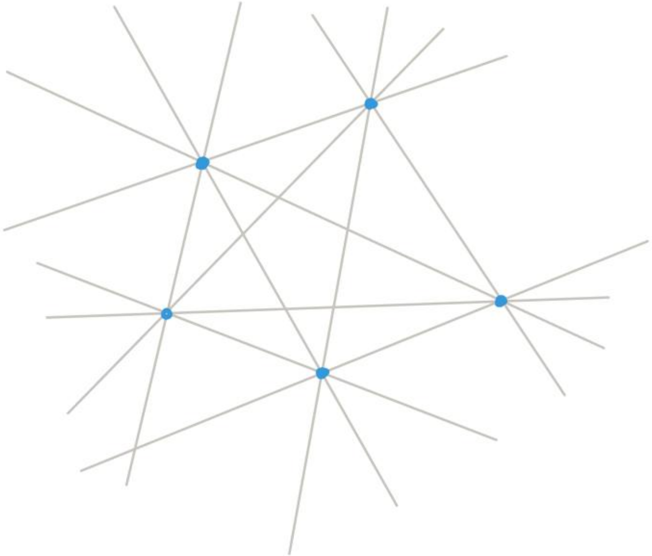
Santos (2012): There exists a 43-dimensional polytope with 86 facets of diameter 44 > 43 = 86 – 43.
Thus, the Hirsch conjecture is *false!*

Davies, Gupta, Racanière, Swirszcz, Wagner, Weber, W.:
“Hopper” algorithm using transformer neural network.
Millions of new counter-examples + smallest known counterexample (19 dimensional).



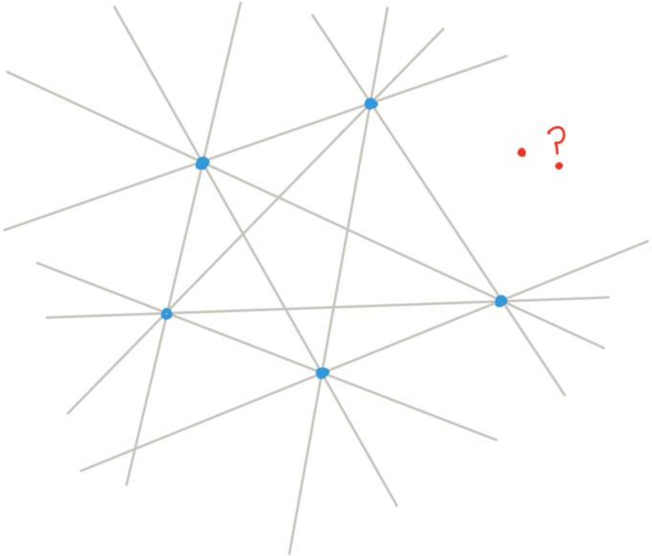
Hopper algorithm

Build up polytope by adding point after point:



Hopper algorithm

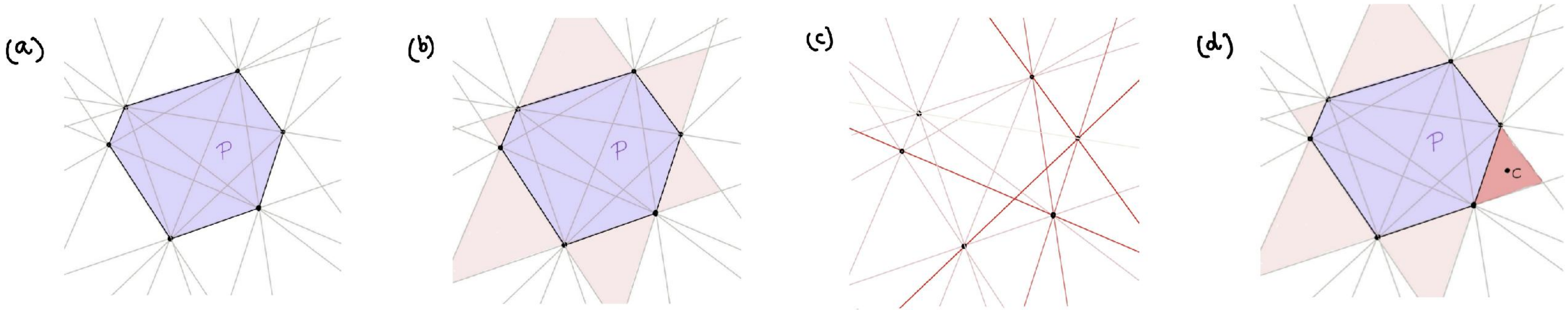
Build up polytope by adding point after point:



Where to add new point?

20 (generic) points in 4 dimensions determine
22 950 110 195 021
possible regions!

Hopper algorithm

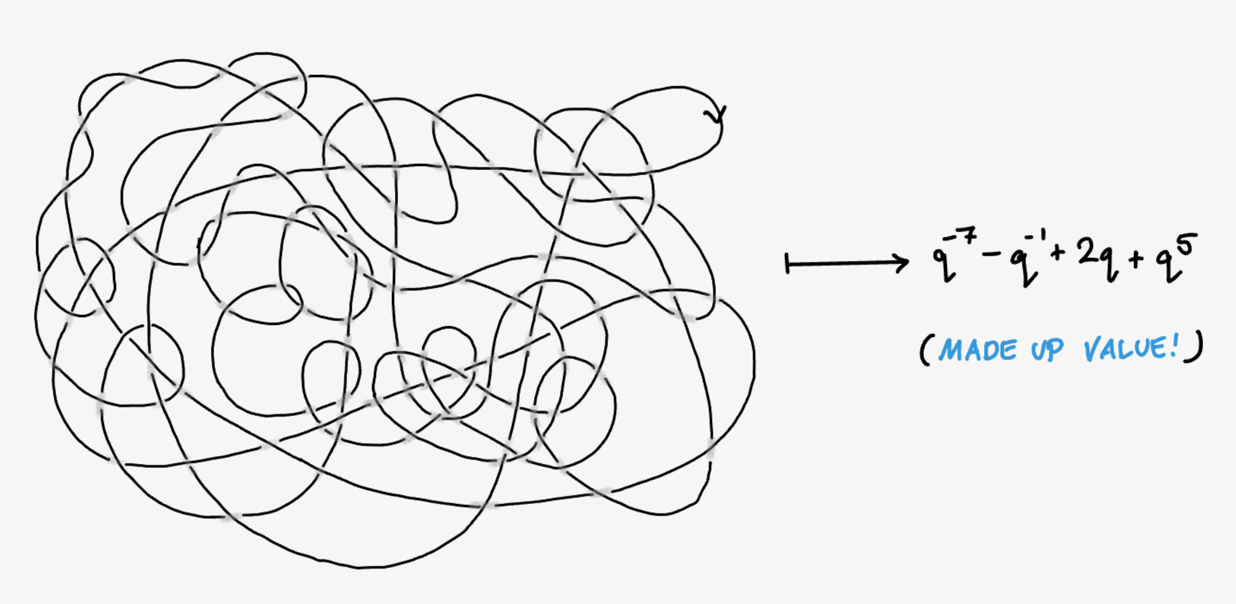


- Start by random sampling. Remember which hyperplanes were involved.
- Transformer neural network tries to predict hyperplanes which will be involved in promising polytopes (with respect to several heuristics).
- Eventually, a tiny fraction (~0.001%) of searches yield a counter-example.

Example 2: Jones unknot conjecture

Does the Jones polynomial detect the unknot?

knot L (oriented) $\longrightarrow$ $V_L \in \mathbb{Z}[q, q^{-1}]$ (Jones polynomial)



Does the Jones polynomial detect the unknot?

Joint work with Charton, Narayanan and Yacobi (building on earlier work with Gibson and Yacobi).

We concentrate on matrix problems: do there exist unexpected relations between matrices. E.g.

$$A = \begin{pmatrix} -q & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 1 & -q^{-1} \end{pmatrix} \quad \text{and} \quad B = \begin{pmatrix} 1-q & -q^{-1} & q^{-1} \\ 1-q^2 & -q^{-1} & 0 \\ 1 & -q^{-1} & 0 \end{pmatrix}$$

Appears to be extremely difficult problem. In [CNWY], several new (non-AI) ideas introduced, which allow us to recover *all known relations, and many new ones*.
However, we don't (yet) find any previously unknown case of non-faithfulness.

The focus here is an interesting part of this work where we use neural networks.

Thought experiment

Imagine you are looking for a rare flower.

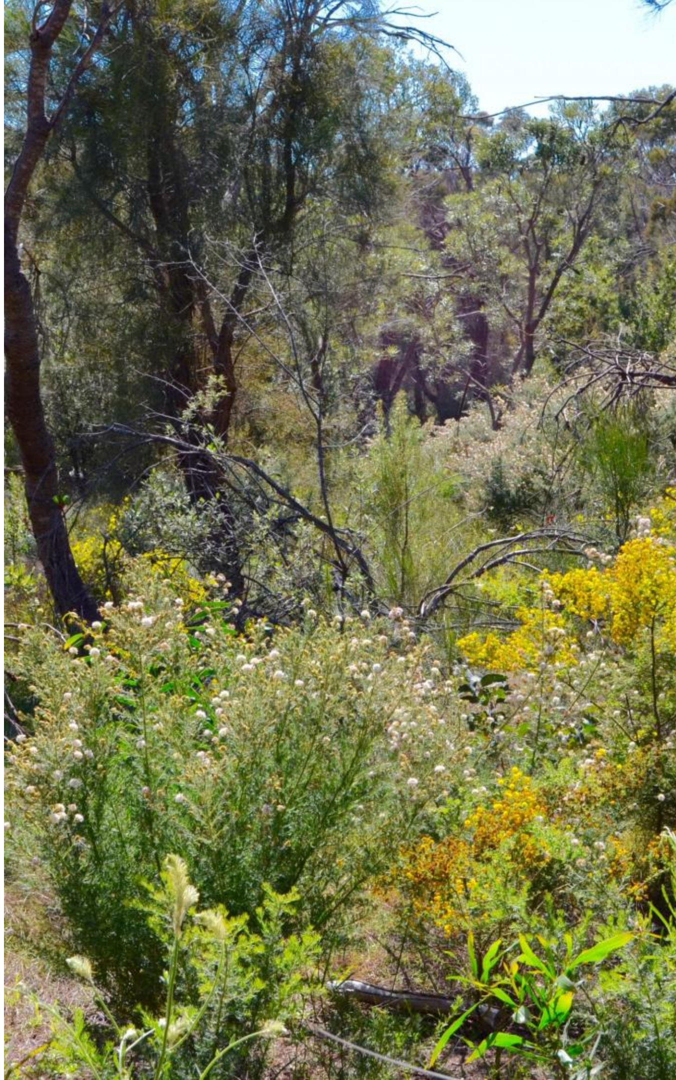
After several days searching you find nothing.

However, you have learnt a lot of other things.

Can these other things be useful in some way?

Areas which seem unusual might be more likely to contain rare species, and in particular the flower you are after.

You might want to concentrate your search on areas of "maximum surprise". That is, areas where your ability to predict your environment is low.



Thought experiment in practice

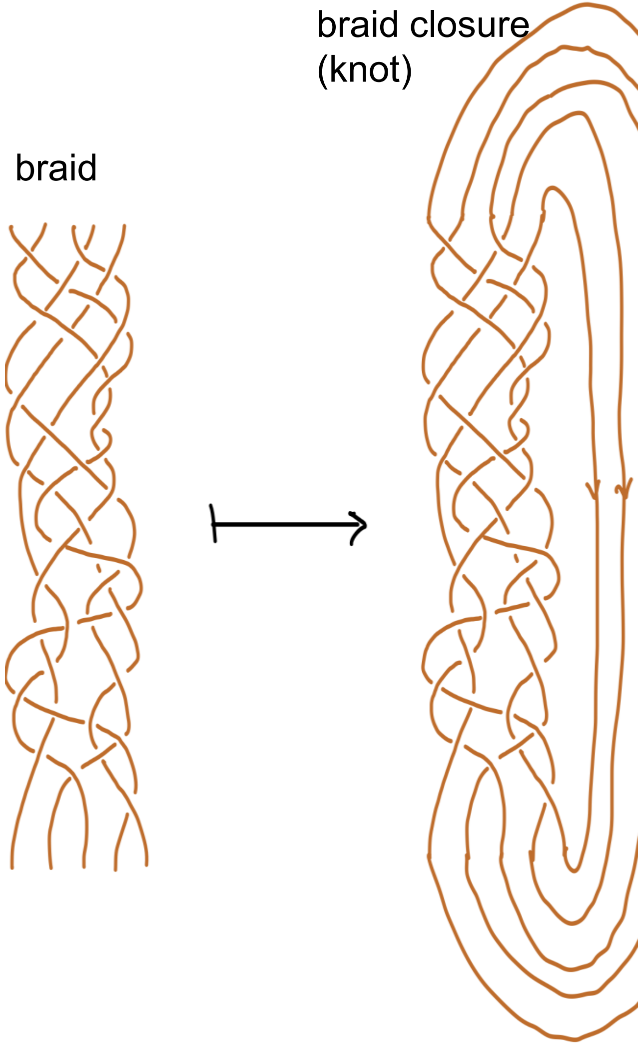
We are looking for a rare knot / braid (a needle in the haystack).

We train a neural network to predict an unrelated quantity (“right descent set of rightmost Garside factor”) on which the neural network can achieve good accuracy.

Neural network achieves good accuracy on this problem.

“Descent set confusion”: l^1 norm between prediction and reality.

Use descent set confusion as a score function,
i.e. braids with high descent set confusion
(i.e. unexpected for neural network)
are more likely to be investigated.

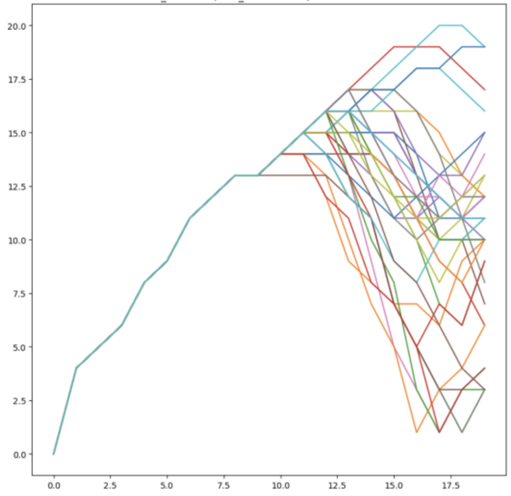


Thought experiment in practice

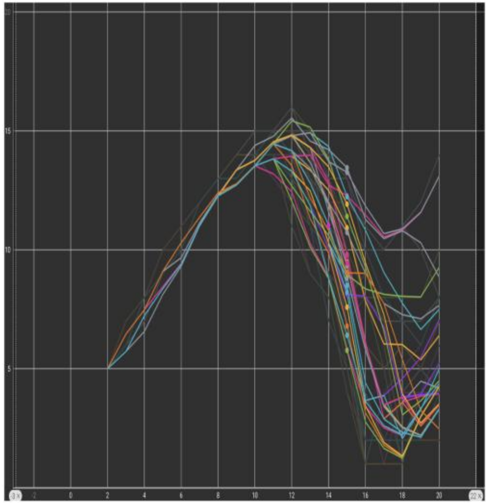
To our delight, this actually works!

On the right we see the results on a toy example. Descent set confusion *dramatically improves* the probability of successful search.

For experts: Here neural network is functioning like a value network in reinforcement learning.



Without neural network, success rate **15%**



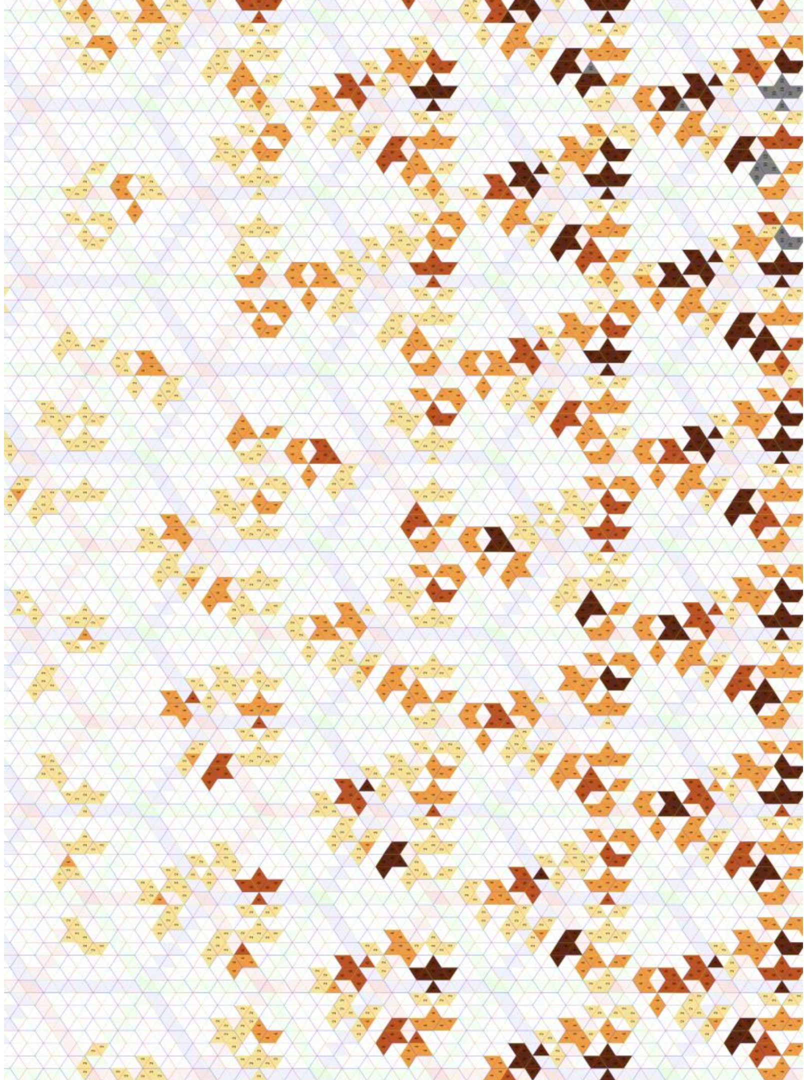
With neural network, success rate 29/39 = **74%**

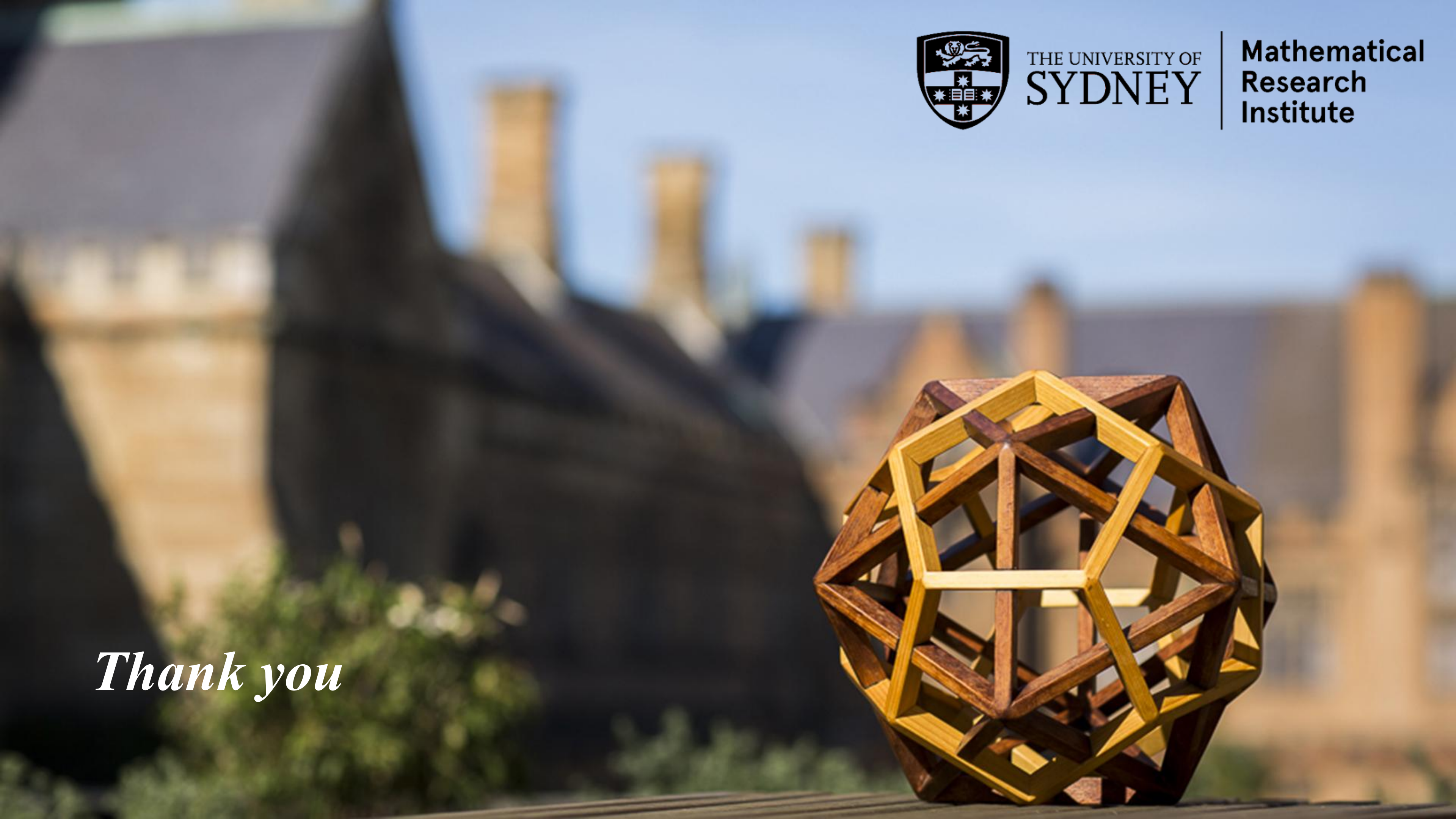
Summary

I have presented three examples where neural networks have helped discover interesting new examples in mathematics.

These are difficult problems, and on no problem do we find a complete solution. However, we do find several new examples and counter-examples.

These techniques are very flexible.
Let’s try them on more problems!





THE UNIVERSITY OF SYDNEY

Mathematical Research Institute

Thank you