CayleyPy: CayleyPy - Artificial intelligence methods for group and graph theories

Alexander Chervov (Paris, Institute Curie)

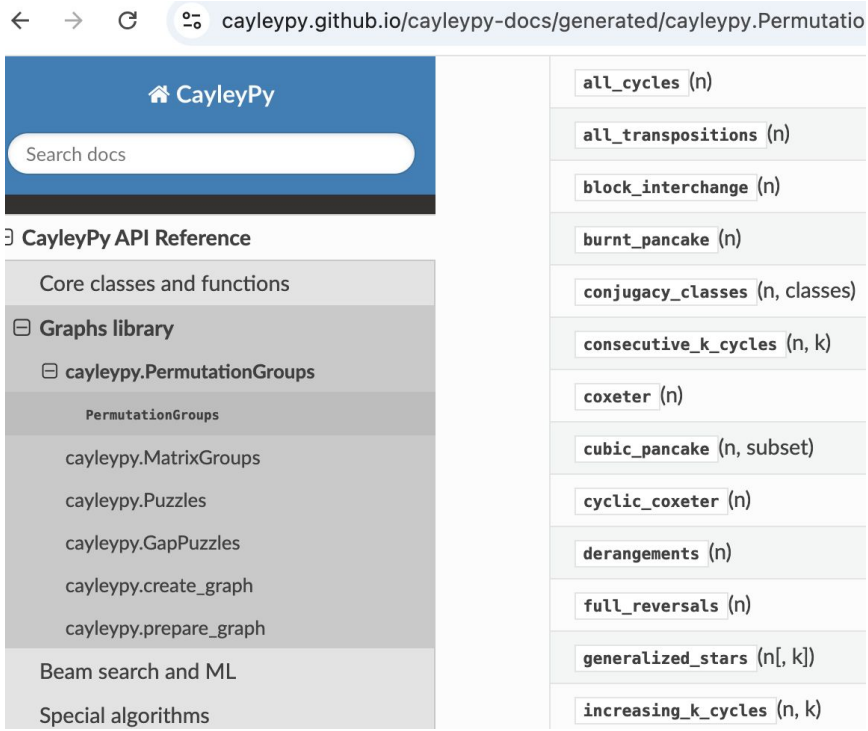
Fall 2025

Crowd-sourcing project 100+ involved:
A. Chervov, D. Fedoriaka, E. Konstantinova, A. Soibelman, A. Smolensky, S.Galkin, F.Petrov, F. Levkovich-Maslyuk, H.Isambert …
CayleyPy-1: Accepted NeurIPS (spotlight) 2025
CayleyPy-2: Accepted in ATMP
CayleyPy-3: arXiv:2509.19162

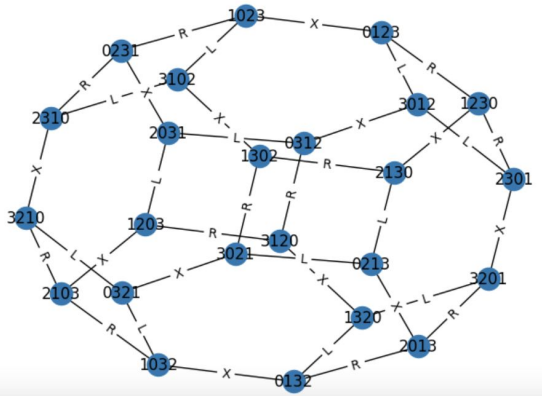
CayleyPy: AI-based open-source Python library to handle GOOGOL size graphs

150+ stars on Github; Documentation; Tutorials; 300 notebooks with experiments

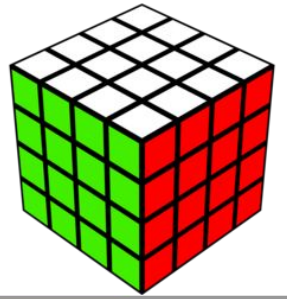
https://cayleypy.github.io/cayleypy-docs/index.html



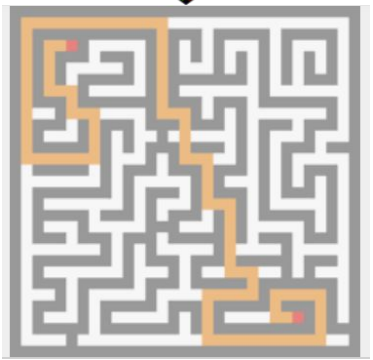

```
import networkx as nx
from cayleypy import CayleyGraph, PermutationGroups

graph = CayleyGraph(PermutationGroups.lrx(4))
result = graph.bfs( return_all_edges=True, return_all_hashes=True)
G = result.to_networkx_graph()

pos = nx.kamada_kawai_layout(G)
nx.draw(G, pos, with_labels=True)
edge_labels = nx.get_edge_attributes(G, 'label')
nx.draw_networkx_edge_labels(G, pos, edge_labels=edge_labels)
None
```



First task - FOUR equivalent ways to explain:



1. **(Fun) Solve a Rubik's cube (any puzzle)**

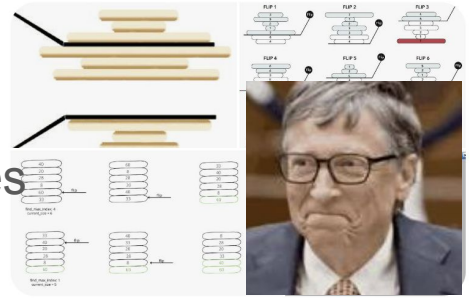


2. **(Math) Decompose group element into product of generators, e.g. decompose a matrix into product of given matrices**

3. **(Graph) Find a path on a graph (Cayley graph)**

Pancake sorting

Sorting algorithm ⋮



4. **(CS) Sorting problem**

Terminology: **Diameter** of graph - largest distance between nodes = God's number = worst case performance of God's algorithm

Results and SOTA

1. C. Sims 1970: Schreier–Sims algorithm (permutation groups)
2. Knuth, Donald E. (1991): randomized version (10^20-10^30)
3. **CayleyPy beats by orders and orders of magnitude (10^50-10^120)**

n	GAP Length	Ideal Length n(n-1)/2	AI Length
9	41	36	36
10	51	45	45
11	65	55	55
12	78	66	66
13	99	78	78
14	111	91	91
15	268	105	105
16	2454	120	120
17	380	136	136
18	20441	153	153
19	3187	171	171
20	217944	190	190
21	-	210	210

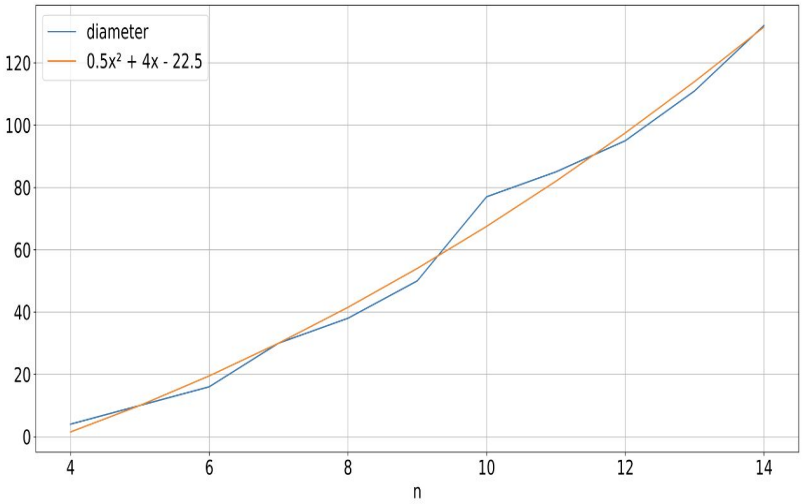
Complexity:

1. 1981 NP-hard: optimal solutions and diameters
2. 1998-...: optimal solution NP-complete for Rubik's cube, etc

Testbed for: automatic provers, "chatGPT", robotics …

L.Babai conjecture - way to Fields medal :)

1. For any generators of S_n , diameter < O(nˆ2) (open: even polynomial bound)
2. CayleyPy 10ˆ2-10ˆ3 times FASTER than GAP, so extensive comp.exp.:
3. New: explicit generators which conjecturaly maximize diameter - follow simple pattern: "square with whiskers" + involutions
4. New conjecture: diameter < nˆ2/2 + 4n



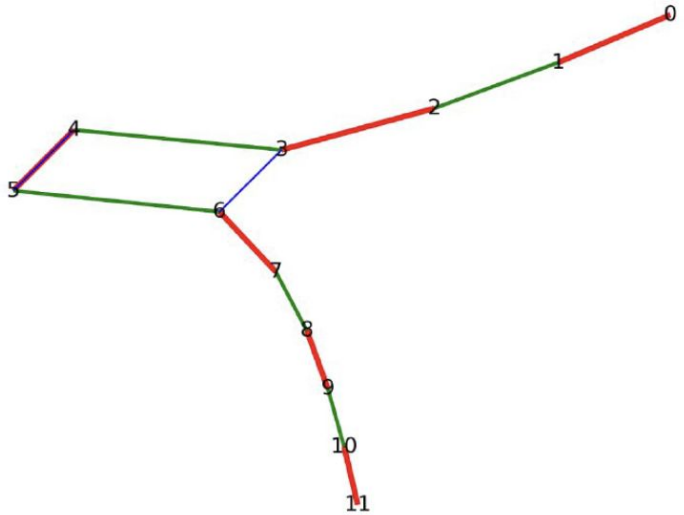


FIGURE 10. Maximal diameters found up to $n \leq 14$ and quadratic fit: $n^2/2 + 4n - 22.5$.

Surprise: Diameters - "findable"(?!) = quasi-polynomials

Example of quasi polynomial: n^2/3 if n = 0 mod 3, (n^2-1)/3 if n=1,-1 mod 3

1. **Conjecture 1: Diameters of S_n are quasi-polynomials (for constructive generators)**
2. **Conjecture 2: Length (word metric) of ANY element is quasi-polynomial (constructive element and constructive generators)**

Example (https://oeis.org/A039745) :

Two generators L - left cyclic shift, X = (1,2)

"father of Soviet cybernetics" V.M.Gluskov 1968 problem:

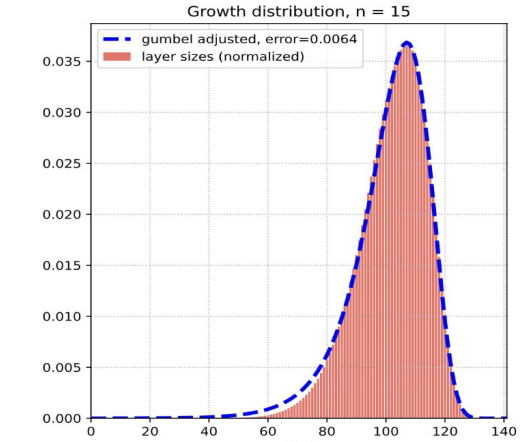
New conjecture: n>4: diameter (3n^2-8n+9)/4, n odd

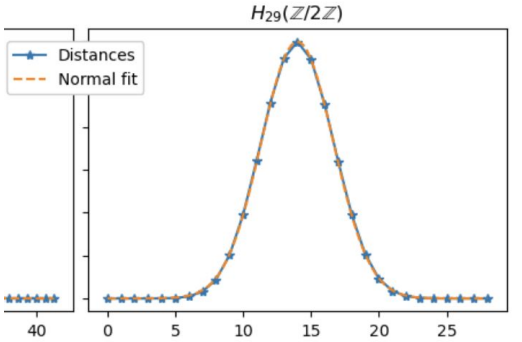
(3n^2-8n+12)/4, n even . Checked n<16

Hundreds conjectures ("life exists beyond diameters")

Black diamond - new conjecture, blue diamond - we can prove

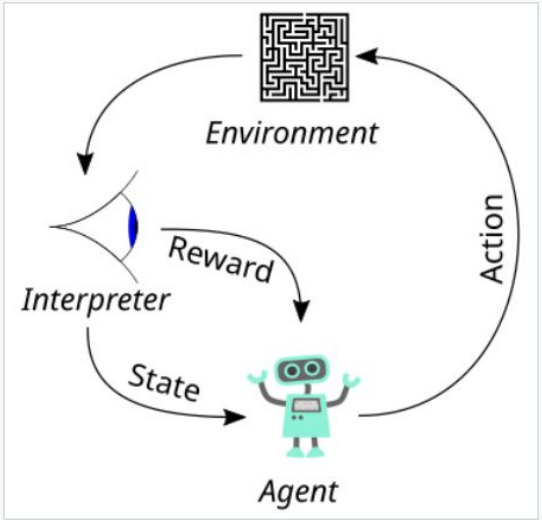
Gene-rators	Gro-up	Dia-meter	Growth							Anti-podes	Algo-rithm	Met-ric	Spec-trum	Mixing Time
			PDF	F-la	Mean	Mode	Var	Skew	Kurt					
Coxeter	S_n	$\frac{n(n-1)}{2}$	Gauss	+	$\frac{n(n-1)}{4}$	$\frac{n(n-1)}{4}$	+	$\to 0$	$\to 0$	1	Bubble	+	?	?
Cyclic Coxeter	S_n	$\left\lfloor \frac{n^2}{4} \right\rfloor$	Gauss◆	?	$0.17(n^2+n+1)$◆	≈Mean◆	?	$\to 0$◆	$\to 0$◆	1|2◆	?	+	Wig◆	?
LRX	S_n	$\frac{n(n-1)}{2}*$	Gumbel◆	?	$\approx 0.38n^2-n$◆	$\approx 0.39n^2-n$◆	?	$\to -0.7$◆	$\to 3.3$◆	◆	◆	?	Uni◆	$> n^3$◆
LX-Glushkov	S_n	$\frac{3n^2-8n+9\|12}{4}$◆	Gumbel◆	Fib/?	$\approx 0.57n^2-2n$◆	$\approx 0.57n^2-1.6n$◆	?	$\to -0.7$◆	$\to 0.5$◆	*	◆	?	?	?
LARX	S_n	$\frac{n^2-2\|5}{2}$◆	?	?	$0.4n^2-0.7n$◆	≈Mean◆	?	?	?	+◆|?	?	?	?	?
LARX+I	S_n	$\frac{n(n+6)-12\|19}{4}$◆	?	?	$\approx \frac{n(n+1)}{4}$◆	≈Mean◆	?	?	?	+◆	?	?	?	?
LSL	S_n	$\frac{n(n-3)}{2}+3$◆	?	?	$\approx 0.4n^2-1.5n$◆	≈Mean◆	?	?	?	+◆	?	?	?	?
LSL+I	S_n	$\frac{n(n+4)}{4}-3\|\frac{17}{4}$◆	?	?	$\approx 0.2n^2$	$\approx 0.2n^2$	?	?	?	?	?	?	?	?
3-cyc	A_n	$\lfloor \frac{n}{2} \rfloor$	?	+−	$\approx D-0.5\|1$◆	$D-0\|1$◆	?	?	?	+	+	+	Int	?
(0ij)	A_n	$\left\lfloor \frac{3(n-1)}{4} \right\rfloor$	?	?	$\approx 0.55n$◆	$\approx Mean$◆	?	?	?	?	?	?	?	?
(01i)	A_n	$\frac{3n-5}{2}+\frac{i^n+(-i)^n}{4}$◆	?	?	$\approx n-2$◆	$n-1$◆	?	?	?	?	+◆	?	?	?
(01i)I	A_n	$\lfloor \frac{3n-6}{2} \rfloor$◆	?	?	$\approx n+1.25\ln n+\ldots$◆	≈Mean◆	◆	◆	◆	◆	◆^{O}	◆	Int◆	nlog(n)◆





RL explained in 1 minute. RL = graph path-finding

Standard (forget it):



The typical framing of a reinforcement learning (RL) scenario: an agent takes actions in an environment, which is interpreted into a reward and a state representation, which are fed back to the agent.

Vocabulary:

States	=	**Nodes of the graph**
Actions	=	**Edges**
Reward(Penalties)	=	**Weights of edges**
Cumulative reward	=	**Length of the path**

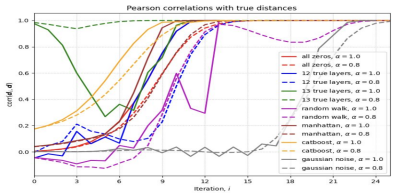
Obviously: d(g,e)=d(g) distance from g to any "e", satisfy:

$$d(g) = 1 + \min_{n:\ neigbours\ of\ g} d(n)$$ Bellman equation

$$d(e) = 0$$ boundary condition

Clearly tropical equation. What is tropical Sandpile ?

Solving Bellam = simple iteration



CayleyPy1 approach - diffusion distance based:

1) **Train set via random walks**: Ran many random walks trajectories starting from identity “id” and store pairs: (g,k): g - node of graph, k - in how many steps it was achieved by random walk. (

   Train set - pairs: (permutation , n_steps)

2) **Model**: Train machine learning model - to predict “k” from “permutation”

3) **Pathfinder (graph search, beam search)**:

   Start from node, take it all neigbours, compute neural net predictions, choose say 10 best (min predictions) nodes, repeat until "id" is found

Neural nets make mistakes: Beam Search to Compensate it

Beam Search = model of research community

Single researcher is exponentially less effective than two (group) of researchers, because "cannot break the wall" (stuck in local minimums)

Pipeline of all RL-like systems (e.g. AlfaGo):

1 Neural Net says where to move

(trained by "playing with itself")

2 Graph search compensates its mistakes (Monte-Carlo Tree Search, Beam Search, A-star, many other global optimization algorithms)

Beam search:

Has a single parameter - size of the beam (size of research group) say 10:

1 Look on all neigbours - choose 10 best

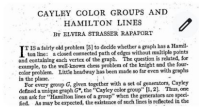
2 Iterate

Surpsie: 10 cannot be compensated by 10 hours of single "researcher", because you stuck in local minimums and only others can researchers can provide breakthroughs to come out

Polymath, Crowd-sourcing - attract 1000+ AI experts to math problem - easy - Kaggle just needs to "press one button"

CayleyPy RapaportM2 - solve math mystery from 1959

Help mathematicians to find the God's number and algorithm - develop method to solve/sort permutations using E.Rapaport transformations


CAYLEY COLOR GROUPS AND HAMILTON LINES

BY ELVIRA STRASSER RAPAPORT

#	Team	Members	Score	Entries	Last
1	Ruslan Grimov		120228357	3	8d
2	Alim Bijiev		153013709	3	13d

CayleyPy Pancake sorting - outperform B.Gates

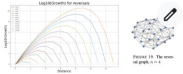
Prefix (pancake) sorting algorithm was proposed by B.Gates in 1979 - can you do better ?

CayleyPy V.M.Glushkov permutations problem

Solve the problem by "father of Soviet cybernetics" V.M.Glushkov (1968)

CayleyPy Reversals - Estimate Evolution Distance

70 million years from human to mice - help to check that is true 😀 Develop reversal sorting algorithms as optimal as possible



Kaggle = "world championship" for AI challenges. (It belongs to Google).

Math problems: optimize 1 number + easy to check correctnes are perfect for Kaggle

Challenges attract THOUSANDS participants. Competition are extremely tough! Is not about money (no chance to get them, (I got twice :)). It is about:

- Adrenaline, fame/glory/honor
- "Reputation" - be "Kaggle grandmaster" - job will find you by itself

Create a challenge - if Kaggle decides "assign reputation" - you got 1000 AI-experts

Biological application - Cayley distance ~= evolutionary distance

Transforming cabbage into turnip: polynomial algorithm for sorting signed permutations by reversals

1999 cited: 1215

Sridhar Hannenhalli, Pavel A. Pevzner

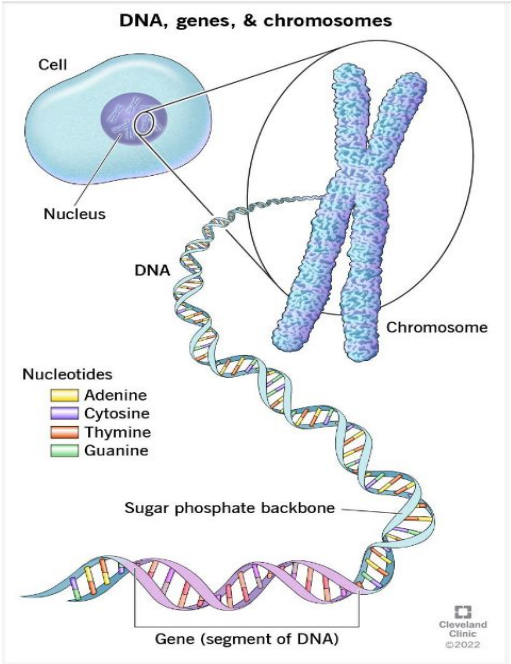
Authors: Sridhar Hannenhalli, Pavel A. Pevzner Authors Info & Claims

Transforming men into mice (polynomial algorithm for genomic distance

5.1.5. *Transposons (block transpositions).*
The generators called somewhat misleadingly in biological literature "transpositions" not to be confused with standard mathematical transpositions, better called "block transpositions", or we will call them just "transposons" - since their biological motivation is to describe "transposons" (movig fragments of genome). They can be described as follows: cutting out a segment and reinserting it elsewhere on the same chromosome (that segment is abstraction of biological transposon):

$$\dots u \underbrace{v \dots w}_{\Delta} x \dots y \dots \longmapsto \dots u\, x \dots y \underbrace{v \dots w}_{\Delta} \dots .$$

They are supported by CayleyPy under the name "transposons".
Conjecture.(OEIS-A065603) Diameter of transposons is $\lceil \frac{n+1}{2} \rceil$ for $n \neq 13, 15$.
The conjecture formulated in [ErikssonErikssonKarlanderSvenssonWästlund01] (section 3 end), based on earlier results in [BafnaPevzner98].



Bill Gates theorem on sorting pancakes

- Gates, W.; Papadimitriou, C. (1979). "Bounds for Sorting by Prefix Reversal". *Discrete Mathematics*. **27**: 47–57. doi:10.1016/0012-365X(79)90068-2.

In 1979, Bill Gates and Christos Papadimitriou[11] gave an upper bound of 5/3n

Diameter (=God's number) - the largest possible distance on the graph = the worst case number of operations for sorting algorithm = the number of moves solve in worst case

Pancake graph - Cayley graph for Sn, with generators: take first k pancakes, swap them and put back.

E.g. k=4: (i1,i2,i3,i4,i5….)->(i4,i3,i2,i1,i5,i6,....)

Pancake Diameter is UNKNOWN - open problem (n>20)

Diameter Rubik cube UNKNOWN n >3 . Etc…

Theorem

If there exists a function $V \in C^1(\mathbb{R}^n;\mathbb{R})$ *such that for all* $x \in \mathbb{R}^n$

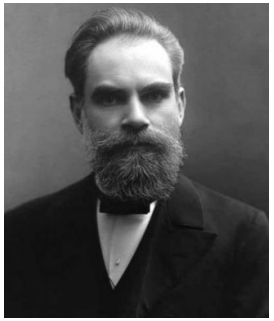
$$V(x) > V(0), \quad \text{and} \quad \nabla V(x) \cdot f(x) \leq 0,$$

and

$$\lim_{\|x\| \to +\infty} V(x) = +\infty,$$

then the system is stable.

A. Lyapunov
(1857-1918)



[Submitted on 10 Oct 2024]

Global Lyapunov functions: a long–standing open problem in mathematics, with symbolic transformers

Alberto Alfarano, François Charton, Amaury Hayat

Neural network architecture: Transformer (∼1000 smaller than GPT-3)

Procedure:

1. Generate a set of systems and associated Lyapunov functions.
2. Encode the examples
3. Train the language model (supervised learning)

[Submitted on 27 Aug 2024 (v1), last revised 11 Feb 2025 (this version, v2)]

What makes math problems hard for reinforcement learning: a case study

Ali Shehper, Anibal M. Medina-Mardones, Lucas Fagan, Bartłomiej Lewandowski, Angus Gruen, Yang Qiu, Piotr Kucharski, Zhenghan Wang, Sergei Gukov

Selman Akbulut and Robion Kirby. "A potential smooth counterexample in dimension 4 to the Poincare conjecture, the Schoenflies conjecture, and the Andrews–Curtis conjecture". *Topology* 24.4 (1985) (cit. on pp. 4, 7).

2. Andrews–Curtis conjecture

The Andrews–Curtis Conjecture concerns the study of *balanced presentations* of the trivial group, i.e. presentations of the trivial group with an equal number of generators and relators. The conjecture proposes that any balanced presentation of the trivial group

$$\langle x_1, \ldots, x_n \mid r_1, \ldots, r_n \rangle$$

can be converted to the trivial presentation

$$\langle x_1, \ldots, x_n \mid x_1, \ldots, x_n \rangle$$

through a series of the following operations known as *AC-moves* [AC65]:

(AC1) Substitute some r_i by $r_i r_j$ for $i \neq j$.
(AC2) Replace some r_i by r_i^{-1}.
(AC3) Change some r_i to $g r_i g^{-1}$ where g is a generator or its inverse.

A notable family of potential counterexamples, denoted AK(n), due to Akbulut and Kirby,

$$\mathrm{AK}(n) = \langle x, y \mid x^n = y^{n+1}, xyx = yxy \rangle, \quad n \geq 3$$

has been open in mathematics for more than four decades [AK85].[3] The length of the presentation AK(n) is $2n+7$, and until now, it was unknown whether this length can even be reduced using AC moves. We find that for every fixed n, AK(n) is AC-equivalent to each element of the following 1-parameter family of presentations parameterized by $k \in \mathbb{Z}$,

$$P(n,k) = \langle x, y \mid y^{n-k-1}x^{-1}yx = xyx^{-1}y^{n-k} \ , \ x = y^{-k}x^{-1}yxy \rangle.$$

When $n \geq 5$, this family contains presentations of length less than $2n+7$. The shortest presentation occurs in the case $k = n-1$, which gives us the following result, proven as Theorem 4 in the main text.

Theorem A. *For every $n \geq 2$, AK(n) is AC-equivalent to the presentation*

$$\langle x, y \mid x^{-1}yx = xyx^{-1}y \ , \ xy^{n-1}x = yxy \rangle,$$

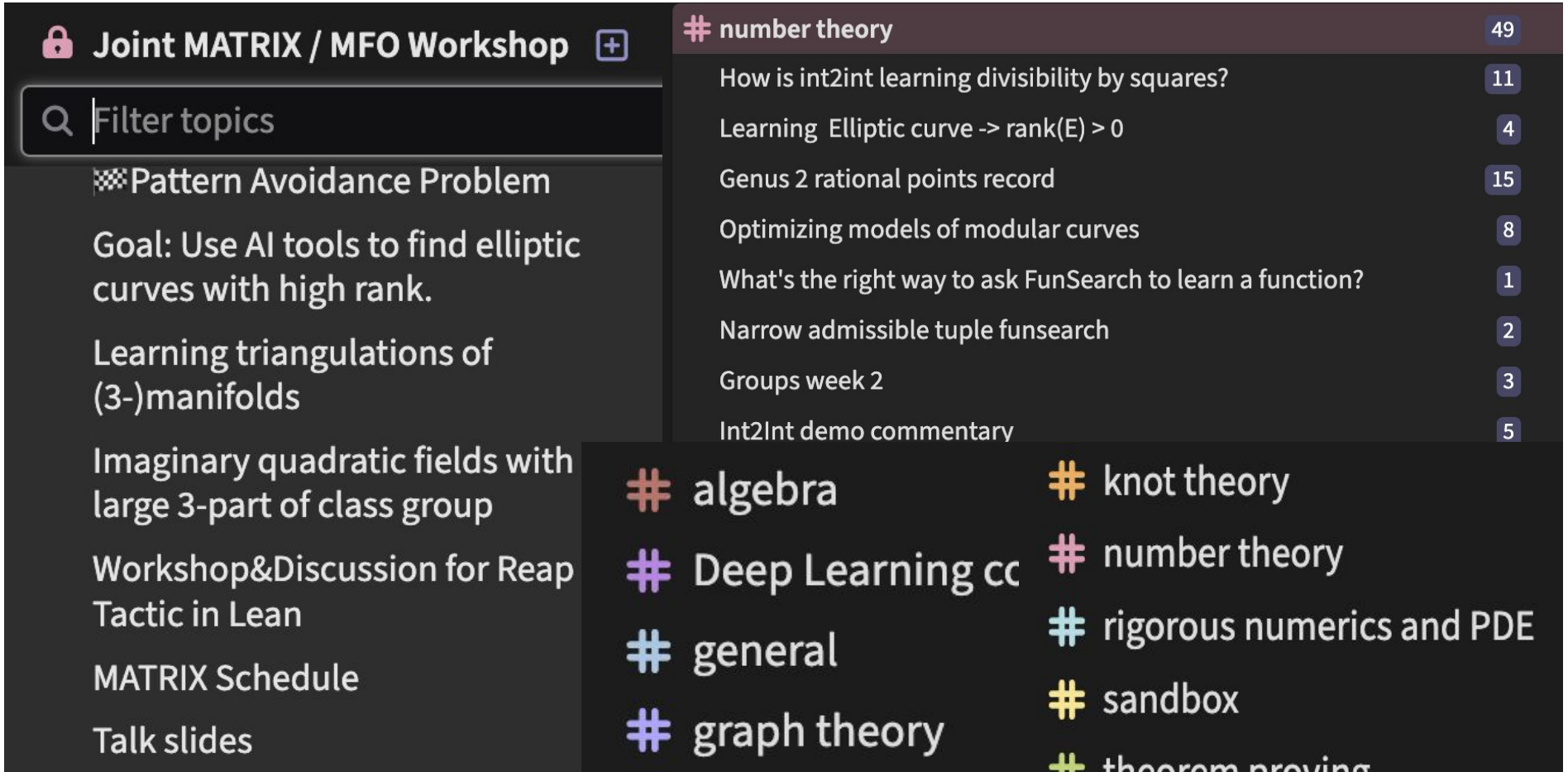
of length $n+11$. This gives a reduction in length of AK(n) for all $n \geq 5$.

Another important family of potential counterexamples, due to Miller and Schupp, is

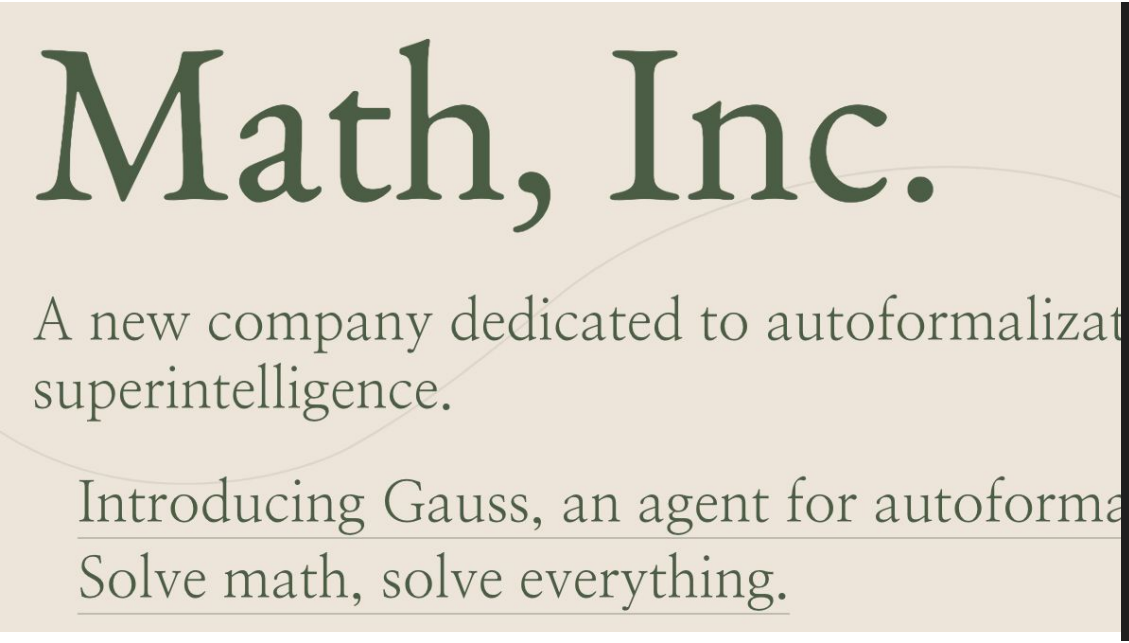
$$\mathrm{MS}(n,w) = \langle x, y \mid x^{-1}y^n x = y^{n+1}, x = w \rangle,$$

where $n > 0$ and w is a word with exponent sum zero on x. It has been open for more than 25 years [MS99].

Merge AI people with math people - find "low laying fruits"



Formalization: Tao, Gowers, Lean, Math.Inc (C.Szegedy) …



Axiom.xyz
https://axiom.xyz › blog › funding

Announcing Axiom's $20 Million Series A Round

Axiom raised $20 million led by Paradigm and Standard Crypto to grow their team and accelerate development of their ZK platform.

CryptoRank
https://cryptorank.io › Funding Rounds

Axiom - Funding Rounds

Discover fundraising information: Funding Rounds, return on investment (ROI), prices of investors, and **funds raised by Axiom**. Review the white paper, team, and ...

FinSMEs
https://www.finsmes.com › USA

Axiom Math Raises $64M in Seed Funding

1 Oct 2025 — Axiom Math, a San Francisco, CA-based startup building self-improving superintelligence, **raised $64M in seed funding**.

LinkedIn · Lisa Chau
4 reactions · 1 month ago

Stanford PhD Student Raises $50M for AI Startup Axiom

"Carina Hong isn't waiting for graduation to take on Wall Street. The **Stanford math PhD student is raising $50 million for her AI startup**, ...

Christian Szegedy

Researcher
Verified email at szegedy.org
Deep learning Formal reasoning

FOLLOW

Cited by		VIEW ALL
	All	Since 2020
Citations	309396	237001
h-index	39	37
i10-index	59	51

Thank you very much !

Revolutionize AI with second super string revolution methods (Mirror, dualities, etc).

Spoiler for the next time:)

Current problem - models need much more data than humans to learn something.

Models ignore structures. (They learn them from data)

("When you fire linguist translation system starts to work better ")

But structures (e.g. grammar) exists.

Is there any clever way to use it to improve learning ?

Hopefully yes and we need string theory methods to achieve it.

Next time hope to tell you more :)