AI-Powered Discovery of Counterexamples in Combinatorics

Yasir Alhasaniyyah, Iurii Beliaev, Thomas Kidane, Kate Newbold, Robert Sun, Kevin Wang

Fan Wei / Math Department / Duke University

Project Overview

Our research leverages cutting-edge machine learning and computational search techniques to tackle long-standing open problems in mathematics. By reframing abstract conjectures as optimization or search problems, we can use AI to explore vast possibility spaces and hunt for elusive **counterexamples**—specific instances that would disprove a long-held belief.

Methodology

We translate abstract mathematical problems into concrete computational tasks that can be solved by AI. Our primary methods include:

- **Gradient-Based Optimization:** Using techniques like gradient descent to navigate the high-dimensional space of possible mathematical objects (e.g., graphs, allocations).
- **Adaptive Monte Carlo Search (AMCS):** Combines a fast, hill-climbing local search for rapid solution refinement with a strategic outer loop that increases search intensity to escape local optima when progress stalls.
- **Reinforcement Learning (RL):** Training AI agents to "build" a counterexample by rewarding actions that move closer to violating a conjecture's conditions.
- **Hybrid (RL + AMCS):** Leverages RL for broad, global exploration to identify a promising region of the solution space. The best solution found by RL is then passed to the fast, exploitative AMCS for high-precision refinement to locate a superior local optimum.

This allows us to automate the search for complex structures that have evaded mathematicians for decades.

Sidorenko's Conjecture

This conjecture claims that in any network, the density of homomorphisms from a bipartite graph H is at least what you'd expect in a random graph.

$$t(H,G) \geq t(K_2,G)^{|E(H)|}$$

Our AI Approach: We treat graphs as matrices and define a loss function to define how "close" is our graph to being a counterexample. We use a range of approaches from Reinforcement Learning and Adaptive Monte Carlo Search to Gradient descent to find a graph that violates the inequality.

Second Neighborhood Conjecture

The conjecture states that for any finite oriented graph D, there exists a vertex v such that:

$$|N^{++}(v)| \geq |N^{+}(v)|$$

Our AI Approach: We turn this into a game for the AI. The goal is to build an oriented graph where *every single node* violates the conjecture. We define a "violation score" and use reinforcement learning to reward the agent for adding or removing connections in a way that increases the extent to which the vertex conditions outlined by the conjecture are violated.

Results

Sidorenko's Conjecture:

- No counterexample was found. However, our **hybrid RL+AMCS method** found a solution that came closer to violating the conjecture than either method alone by orders of magnitude.
- This validates the strategy of using RL for broad exploration and AMCS for precise refinement.

Method	Conjecture Gap	Time (s)
RL Alone	5.70×10^{-8}	143.18
AMCS Alone	8.23×10^{-11}	328.58
Hybrid	$\mathbf{6.08 \times 10^{-15}}$	**153.08**

Table: Hybrid search finds superior solutions

Second Neighborhood Conjecture:

- Initial searches tended to evolve graphs toward structures with a single "sink" vertex, which trivially satisfies the conjecture and suggests it is quite robust. To counteract this, the minimum out-degree of any vertex in a generated graph had to be at least 7.
- Best performing graphs under reward conditions tended to:
  - Be dense in the core
  - Be sparse in the periphery
  - Contain at least one vertex with a very large second neighborhood

Tools:

- Implemented custom graph homomorphism counting library leveraging dynamic programming and tree decomposition; extended for graphon representations; runs in $\mathcal{O}\big(\mathrm{poly}\big(|V(H)|\big)\,|V(G)|^{\mathrm{tw}(H)+1}\big)$ complexity
- Trained graph attention network (GAT) to model homomorphism density distribution for fixed target graph; predicts homomorphism density with 1.06% maximum and 0.21% average error

Conclusion & Future Work

Conclusion

Our research demonstrates that reframing mathematical conjectures as optimization problems creates a powerful avenue for exploration using AI. We developed a novel hybrid algorithm combining Reinforcement Learning and Adaptive Monte Carlo Search, which proved its effectiveness in comparison to its predecessors on the Sidorenko conjecture. Simultaneously, we found that for the Second Neighborhood conjecture evolutionary search consistently evolves graphs towards structures that are closer to counterexamples.

Future Work

Our next steps are focused on enhancing our methods and expanding their application:

- **Integrating Advanced Models:** We will incorporate more sophisticated Graph Neural Networks (GCN, GAT) to improve our model's expressivity and use transfer learning to scale our search to larger, more complex graphs.
- **Exploring New Conjectures:** We will apply our validated hybrid search framework to investigate other unresolved conjectures in graph theory and combinatorics.

References

1. Sidorenko, A. F. (1993). A correlation inequality for bipartite graphs. *Graphs and Combinatorics, 9*(1), 201-204.
2. Seymour, P. (Conjecture). See "The second neighborhood problem" by DeVos, et al.

Contact Information

- Web: https://math.duke.edu
- Email: math@duke.edu

Duke | Math+