Inequality Ranking and Inference System (IRIS): Giving Mathematical Conjectures Numerical Value

Randy Davila [1 2] **Jesús A. De Loera** [3] **Jillian Eddy** [3] **Ethan Fang** [4] **Junwei Lu** [5] **Zini Yang** [4]

Abstract

We introduce **IRIS**, a geometric and heuristic-based scoring system for evaluating mathematical conjectures and theorems expressed as linear inequalities over numerical invariants. The **IRIS** score reflects multiple dimensions of significance—including sharpness, diversity, difficulty, and novelty—and enables the principled ranking of conjectures by their structural importance. As a tool for fully automated discovery, **IRIS** supports the generation and prioritization of high-value conjectures. We demonstrate its utility through case studies in convex geometry and graph theory, showing that **IRIS** can assist in both rediscovery of known results and proposal of novel, nontrivial conjectures.

1. Introduction

The formulation of mathematical conjectures has long been seen as a deeply creative act—emerging from intuition, pattern recognition, and structural insight. From Ramanujan's identities to the unsolved problems scattered across every domain of mathematics, conjecture-making captures the conceptual leaps often viewed as uniquely human.

This view, however, is evolving. Over the past several decades, automated systems have begun to assist in the generation of mathematical conjectures, often producing results that inform or inspire new theorems—a field of research known as *automated conjecturing*. Among the earliest and most influential was Fajtlowicz's *Graffiti* (Fajtlowicz, 1987; 1990), whose conjectures have led to well over 60 publications in mathematical journals. Other systems have approached conjecturing through geometric and symbolic means. These include *GraPHedron* (Mélot, 2008), which framed conjecture discovery as a convex hull problem, and its successor program *PHOEG* (Devillez et al., 2019). Still others invoke optimization techniques, such as *TxGraffiti* (Davila, 2024a) and the *Optimist* agent (Davila, 2024d), or incorporate neural methods (Davies et al., 2021).

In this paper, we introduce **IRIS**: the **I**nequality **R**anking and **I**nference **S**ystem. **IRIS** assigns each inequality a structured feature vector that encodes geometric and heuristic signals—such as boundary tightness, dimensional diversity, and directional novelty relative to previously accepted conjectures. While **IRIS** is not itself a fully automated conjecturing system, its scoring framework can be applied retroactively to inequalities proposed by systems like *GraPHedron*, *TxGraffiti*, or the *Optimist*, or used prospectively within new automated pipelines as a scoring module. We demonstrate the system's effectiveness through case studies in graph theory and convex geometry, showing that it can recover classical results and propose new conjectures, some of which parallel longstanding open problems in discrete mathematics.

2. Related Automated Conjecturing Systems

We begin by surveying key automated systems that exemplify distinct approaches to mathematical conjecturing. These include symbolic heuristic systems (e.g., *Graffiti*), geometric approaches grounded in polyhedral reasoning (e.g., *GraPHedron*, *PHOEG*), hybrid systems that incorporate optimization techniques (e.g., *TxGraffiti*, *Optimist*), and recent efforts that integrate neural methods with symbolic reasoning (e.g., (Davies et al., 2021)). Although **IRIS** is not itself a conjecturing engine, it is designed to evaluate and score the types of inequalities produced by such systems.

2.1. *Graffiti*

Siemion Fajtlowicz's *Graffiti* (Fajtlowicz, 1987; 1990), developed in the 1980s, is arguably the most influential of all

*Equal contribution [1]Department of Computational Applied Mathematics & Operations Research, Rice University, Houston TX, United States [2]Research and Development, RelationalAI, Berkeley CA, United States [3]Department of Applied Mathematics, University of California – Davis, Davis CA, United States [4]Department of Biostatistics & Bioinformatics, Duke University, Durham NC, United States [5]Department of Biostatistics, T.H. Chan School of Public Health, Harvard University, Cambridge MA, United States. Correspondence to: Jillian Eddy <jeddy@ucdavis.edu>.

Proceedings of the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

automated conjecturing systems. Its conjectures have led to over 60 publications by leading mathematicians. At its core, *Graffiti* operated on a collection of a few hundred graphs. When asked to conjecture, it computed the values of dozens of graph invariants and searched for symbolic inequalities among them, retaining only those that satisfied two key conditions—together known as the *Dalmatian heuristic* (Larson & Cleemput, 2016). The first is the *truth test*: an inequality must hold for all known examples in the dataset. The second is the *significance test*: it must offer a strictly better bound than existing conjectures on at least one example. This filtering process enforced what Fajtlowicz called the *Principle of the Strongest Conjecture*, selecting statements that were both empirically valid and non-redundant.

Over time, *Graffiti* generated hundreds of nontrivial and often deep conjectures, particularly in areas such as chemical graph theory (Fowler, 1997; Fowler et al., 1998; 1999; Fajtlowicz & Larson, 2003; Fajtlowicz et al., 2005; Fajtlowicz, 2005) and graph distances (Fajtlowicz, 1988; Chung, 1988). Its influence extended beyond published theorems—it also played a transformative role in undergraduate research and collaborative discovery (Pepper, 2001). For a detailed retrospective on the system's development and mathematical impact, see (DeLaViña, 2005).

2.2. *GraPHedron* and *PHOEG*

A distinct approach to automated conjecturing emerged in the early 2000's from geometric reasoning, particularly through the *GraPHedron* system (Mélot, 2003). Rather than relying on symbolic heuristics, *GraPHedron* framed conjecture generation as a problem in convex geometry. Graphs were embedded as points in a high-dimensional Euclidean space, where each coordinate represented a real-valued invariant such as order, size, or diameter.

The core insight was both elegant and powerful: linear inequalities among invariants correspond to the supporting hyperplanes of the convex hull of these points. By computing the facets of this polytope, the system extracted inequalities that were automatically satisfied by all graphs in the dataset—conjectures justified directly by geometric structure. Each graph was thus identified with a point in $\mathbb{R}^d$, where the coordinates correspond to a fixed list of invariants $i_1, \dots, i_d$, and $i_j(G)$ denotes the value of the j-th invariant on graph G. The resulting facet inequalities were interpreted as candidate conjectures of the form

$$a_1 i_1(G) + \cdots + a_d i_d(G) \le b.$$

The strength of *GraPHedron* lies in its exhaustive and unbiased character: all sharp linear inequalities derivable from the dataset are identified. However, its primary limitation is computational. Convex hull algorithms scale poorly with both the number of graphs and the number of invariants, restricting practical usage to relatively small graph classes and low-dimensional invariant spaces.

Building on these ideas, *PHOEG* (Polyhedral Help for Obtaining Extremal Graphs) extends *GraPHedron*'s methodology to larger datasets and improved scalability (Devillez et al., 2019). *PHOEG* treats extremal graph theory as a structured data mining problem, combining a relational database of graphs and precomputed invariants with fast convex hull approximations, facet discovery heuristics, and interactive query tools for retrieving extremal examples and conjectured inequalities.

2.3. *TxGraffiti*

TxGraffiti (Davila, 2024a), developed in the mid-2010s, approaches conjecture generation as a structured optimization process over invariant data. Given a dataset of mathematical objects annotated with numerical invariants and Boolean properties, the system selects a target invariant and constructs upper or lower bounds by solving a sequence of linear programming problems. Each optimization model is restricted to a Boolean-defined subclass of graphs (e.g., connected graphs, connected cubic graphs, connected regular graphs) and produces candidate inequalities that hold across all known instances. These are then filtered using a suite of heuristics inspired by *Graffiti*, including truth, sharpness, generality, and significance tests (Davila, 2024d).

Beyond its research contributions—see, for instance, (Caro et al., 2022; Davila & Henning, 2019; 2021; Brimkov et al., 2024; Davila & Henning, 2020; Davila, 2024c; Schuerger et al., 2024)—*TxGraffiti* also emphasizes accessibility. A public-facing interface (Davila, 2024b) allows users to explore conjectures across standard and custom graph families, lowering the barrier to engagement for both students and researchers.

Although originally developed for graph-theoretic discovery, the mechanisms of *TxGraffiti* are not domain-specific. Like *Graffiti*, the system operates purely on invariant data and Boolean filters, independent of the underlying mathematical objects. As demonstrated in (Davila, 2024a), it has been applied to conjecture over positive integers, matrix-derived invariants, and even non-mathematical datasets such as the wine quality dataset from `scikit-learn`.

2.4. *DeepMind Approach*

In a recent contribution to mathematical discovery, researchers at DeepMind applied machine learning techniques to identify novel patterns among invariants of knots and posets (Davies et al., 2021). Their system was trained on a dataset of over one million knots, each annotated with geometric and algebraic features such as hyperbolic volume, injectivity radius, and signature. Remarkably, the model uncovered an unexpected correlation between the algebraic invariant known as the *signature* and a newly defined geometric quantity called the *natural slope*.

This led to a conjecture bounding their difference:

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

where $\sigma(K)$ denotes the knot signature, $\mathrm{vol}(K)$ its hyperbolic volume, and c_1, c_2 are constants. The conjecture was later proven and refined into a rigorous theorem by human mathematicians, establishing a precise inequality involving signature, slope, volume, and injectivity radius.

This work stands out for integrating large-scale neural models into the discovery pipeline and for catalyzing a human proof of a previously unknown inequality. While earlier systems such as *Graffiti*, *Graffiti.pc* (DeLaViña et al., 2024), and *TxGraffiti* have also led to provable theorems—many in geometric domains—the DeepMind approach illustrates how machine learning can complement traditional methods in modern mathematical research.

3. Problem Setup

Let $\mathcal{O} = \{O_1, O_2, \ldots, O_N\}$ denote a finite collection of mathematical objects—such as graphs, polytopes, or other combinatorial or geometric structures. Each object $O_i \in \mathcal{O}$ is associated with a feature vector $x^{(i)} \in \mathbb{R}^k$, where each coordinate represents a real-valued invariant computed on O_i. Collectively, these vectors form a dataset $X = \{x^{(1)}, x^{(2)}, \ldots, x^{(N)}\} \subset \mathbb{R}^k$, which we interpret as a table of invariant data: each row corresponds to an object, and each column to a specific invariant.

We define a *conjecture* over this dataset as a symbolic inequality of the form

$$a^\top x \leq b,$$

where $a \in \mathbb{R}^k$ is a vector of coefficients and $b \in \mathbb{R}$ is a scalar threshold. A conjecture is said to be *valid* under a Boolean predicate $p \colon \mathcal{O} \to \{\texttt{True}, \texttt{False}\}$ if it holds for all vectors $x^{(i)} \in X$ such that $p(O_i) = \texttt{True}$. This allows us to reason over structural subclasses, such as regular graphs, fullerene polytopes, or trees, without requiring the inequality to hold universally.

To evaluate conjectured inequalities, we adopt a geometric perspective that is agnostic to how the inequalities are produced. Specifically, we consider a blackbox system that proposes inequalities of the form $a^\top x \leq b$, intended to hold over the relevant subset of the dataset defined by a hypothesis p. Such inequalities may originate from symbolic reasoning, convex hull methods, optimization pipelines, or neural models, and are the candidate conjectures we consider.

Each valid inequality can be interpreted as defining a closed halfspace that contains the filtered point cloud $X_p = \{x^{(i)} \in X : p(O_i) = \texttt{True}\} \subset \mathbb{R}^k$. A collection $\mathcal{S} \subseteq \mathcal{H}$ of such inequalities defines a polyhedral outer approximation of the data:

$$\mathcal{P} = \bigcap_{(a,b)\in\mathcal{S}} \left\{ x \in \mathbb{R}^k : a^\top x \leq b \right\},$$

where $\mathcal{H}$ denotes the full pool of candidate inequalities. Each element of $\mathcal{S}$ is treated as a potential conjecture describing the structure of the invariant data.

This geometric framing motivates many of the scoring and ranking techniques in **`IRIS`**. Rather than judging a conjecture by its origin, we assess its mathematical significance based on how it interacts with the ambient geometry of the dataset.

4. `IRIS`

`IRIS`, the **I**nequality **R**anking and **I**nference **S**ystem, assigns numerical scores to valid inequalities interpreted as conjectures over a dataset of invariant vectors. Given a valid inequality $a^\top x^{(i)} \leq b$ for all $x^{(i)} \in X$, we define the *touch set* of a conjecture $h = (a, b)$ as the set of data points that lie approximately on the boundary defined by the inequality:

$$\mathcal{T}(h; X) = \left\{ x \in X : |a^\top x - b| < \varepsilon \right\}.$$

We then define the *touch matrix* $\mathcal{M}(h; X) \in \mathbb{R}^{k \times t}$ as the matrix whose columns are the vectors in $\mathcal{T}(h; X)$, where $t = |\mathcal{T}(h; X)|$. The **`IRIS`** framework evaluates each conjecture using a collection of geometric and heuristic measures derived from this matrix and the ambient dataset X.

Normalized Touch $T(h; X)$: the fraction of data points that lie approximately on the boundary defined by h,

$$T(h; X) = \frac{1}{N} |\mathcal{T}(h; X)|.$$

This score captures *sharpness*: inequalities that are tight for many instances are more likely to reflect extremal structural boundaries in the data.

Normalized Rank $R(h; X)$: the rank of the touch matrix, normalized by the number of its columns,

$$R(h; X) = \frac{\mathrm{rank}(\mathcal{M}(h; X))}{d(\mathcal{M}(h; X))},$$

where $d(\mathcal{M}(h; X))$ is the number of columns of the touch matrix. The rank reflects the dimension of the affine subspace spanned by the touch points. Higher normalized rank suggests the touch set is well-distributed and nearly full-dimensional in the space of selected invariants.

Before defining our next measure, recall that the *diameter* of a finite set $A \subset \mathbb{R}^k$ is given by

$$\mathrm{diam}(A) = \max_{x,y \in A} \|x - y\|_2.$$

Normalized Diameter $L(h; X)$: the diameter of the touch set, normalized via the diameter of the full dataset,

$$L(h; X) = \frac{\mathrm{diam}(\mathcal{T}(h; X))}{\mathrm{diam}(X)}.$$

This score reflects the geometric extent of the conjecture's active region. It approximates the length of an exposed face defined by the inequality, normalized by the overall spread of the data cloud.

Each conjecture may also be assigned scores reflecting its relationship to the set of inequalities to which it belongs. Let $\mathcal{S}$ denote the current set of accepted conjectures. Define the Dalmatian set of $h \in \mathcal{S}$ as follows:

$$\mathcal{D}(h; S, X) = \{x \in X : b - a^\top x \leq \min_{(a',b') \in \mathcal{S}} (b' - {a'}^\top x)\}$$

The following scores give numerical ranking to each conjecture in a set.

Binary Significance $\Delta(h; \mathcal{S})$: a binary indicator that equals one if the conjecture h provides the tightest (or equally tightest) bound for at least one data point in X compared to all other conjectures in $\mathcal{S}$. Formally,

$$\Delta(h; \mathcal{S}, X) = \mathbb{I}\left[\mathcal{D}(h; S, X) \neq \emptyset\right],$$

where $\mathbb{I}[\cdot]$ denotes the indicator function. Note, only accepting conjectures with this score nonzero is equivalent to the *Dalmatian* heuristic of *Graffiti*.

Rational Significance $\sigma(h; \mathcal{S})$: a soft measure indicating the fraction of data points for which h is the most significant conjecture in terms of slack:

$$\sigma(h; \mathcal{S}, X) = \frac{1}{N} |\mathcal{D}(h; S, X)|,$$

where $\mathbb{I}[\cdot]$ denotes the indicator function.

Proximity $P(h; \mathcal{S}, X)$: the normalized inverse slack between the inequality and the dataset,

$$\bar{\delta}(h) = \frac{1}{N} \sum_{x \in X} \left(b - a^\top x\right),$$

$$P(h; \mathcal{S}) = 1 - \frac{\bar{\delta}(h)}{\max\limits_{h' \in \mathcal{S} \cup \{h\}} \bar{\delta}(h')}.$$

This term favors conjectures that lie closer to the boundary of the data cloud, indicating a more difficult or tight inequality.

Angular Novelty $N(h; \mathcal{S})$: the minimum angular distance between the normal vector a and those of previously accepted conjectures,

$$N(h; \mathcal{S}) = \min_{(a',b') \in \mathcal{S}} \left\| \frac{a}{\|a\|} - \frac{a'}{\|a'\|} \right\|_2.$$

This encourages directional diversity in the space of conjectures, promoting new inequalities that cut through invariant space in distinct ways.

Together, these scoring metrics define **`IRIS`**. This set of metrics is modular and can be applied in conjunction with any system that produces linear conjectures. However, it is important to emphasize that **`IRIS`** is not itself an automated conjecturing system. It does not generate inequalities or explore mathematical structures autonomously. Instead, it serves as a flexible evaluation layer: a scoring framework that allows existing systems to assess, filter, and rank the conjectures they produce.

$$\phi(h; X) = \begin{bmatrix} T(h; X) \\ R(h; X) \\ L(h; X) \end{bmatrix} \qquad \psi(h; \mathcal{S}, X) = \begin{bmatrix} P(h; \mathcal{S}, X) \\ N(h; \mathcal{S}, X) \\ \Delta(h; \mathcal{S}, X) \\ \sigma(h; \mathcal{S}, X) \end{bmatrix}$$

Using these vectors derived from the scores of **`IRIS`**, we now define a simple automated conjecturing system. Specifically, given a threshold $\theta \geq 0$, a budget $R > 0$, and weight vectors $\boldsymbol{\lambda} = [\lambda_1, \lambda_2, \lambda_3]^\top$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3, \beta_4]^\top$, we process a stream of valid linear inequalities produced by a blackbox system (e.g., the convex hull of invariant vectors). A candidate conjecture h is accepted into the working set $\mathcal{S}$ of conjectures, if its composite score $\boldsymbol{\beta}^\top \psi(h; \mathcal{S}, X)$ exceeds the threshold θ. After each addition, retain only the top R inequalities in $\mathcal{S}$ based on the composite score $\boldsymbol{\lambda}^\top \phi(h; X)$.

Algorithm 1 NUEVAMIRADA: Filtering Conjectures with **`IRIS`** Scoring

Require: Threshold $\theta \geq 0$, Budget $R > 0$, Weights $\boldsymbol{\lambda} \in \mathbb{R}^3$, $\boldsymbol{\beta} \in \mathbb{R}^4$, Data X

```
Initialize working set S ← ∅
while streaming conjecture h from blackbox generator do
    if β^⊤ ψ(h; S, X) ≥ θ then
        S ← S ∪ {h}
        Sort S by structural score λ^⊤ φ(h; X) in descending order
        Retain top R elements of S
    end if
end while
```

5. Two Case Studies: Discovery and Rediscovery in Graph Theory and Geometry

Our goal is to assess whether **IRIS** scoring—when applied as a ranking and filtering layer within NUEVAMIRADA—can rediscover known relationships, refine structural bounds, or highlight previously overlooked inequalities. In the results that follow, we report on the system's performance in rediscovering classical inequalities involving domination, independence, and zero forcing, as well as its ability to identify meaningful structure in polyhedral face vectors.

Each dataset used in our experiments is structured as a table: rows correspond to individual mathematical objects (graphs or polytopes), and columns represent computed properties—both numerical invariants (e.g., domination number, number of triangular faces) and Boolean attributes (e.g., regular, bipartite, fullerene). These tables define the ambient data space in which **IRIS** operates.

For simplicity, we fix the hyperparameters of NUEVAMIRADA as follows: threshold $\theta = 0.25$, budget $R = 100$, structural weight vector $\boldsymbol{\lambda} = [1, 1, 1]^\top$, and discovery score weight vector $\boldsymbol{\beta} = [1, 1, 1, 1]^\top$. Users seeking different types of conjectures may wish to experiment with alternative settings.

Geometric dataset. The first dataset consists of 496 three-dimensional convex polytopes. By a classical theorem of Steinitz, every convex 3-polytope corresponds to a 3-connected planar 3-regular graph. Moreover, by a theorem of Whitney (see Corollary on page 246 of (Whitney, 1933)), each such graph admits a unique convex embedding. For each polytope, we extract its p-vector, where p_k denotes the number of k-gonal faces in the embedding. These p-vectors define the invariant space in which **IRIS**-based filtering is applied.

Graph theory dataset. The second dataset consists of invariants computed on the 335 simple connected graphs used by the *TxGraffiti* conjecturing website (Davila, 2024b). As our main comparison in graph theory, we focus on connected graphs with maximum degree at most three—a rich and well-studied class for which *TxGraffiti* has already produced published conjectures. This setting allows for meaningful comparison against a system whose conjectures have already led to mathematical results. In particular, we test whether NUEVAMIRADA, equipped with **IRIS** scoring, can rediscover or reimagine conjectures previously posed by *TxGraffiti*.

5.1. On the p-vectors of simple convex 3-polytopes

The p-vector of a simple, 3-dimensional convex polytope P is defined as $p(P) = (p_3, p_4, \ldots, p_m)$ where p_k represents the number of k-gonal 2-dimensional faces of P, and m is the largest k such that $p_k > 0$. In the particular case of a simple 3-polytope, we can reformulate Euler's relation $f_0 - f_1 + f_2 = 2$ in terms of the entries of the p-vector:

$$\sum_{k\geq 3}(6-k)p_k = 12. \tag{1}$$

Eberhard showed that a sequence $(p_3, p_4, p_5, p_7, \ldots)$ satisfying Equation 1 is necessarily realizable as a simple convex 3-polytope for some value(s) of p_6 (Grünbaum, 2003). Today, the problem of classifying *for which* value(s) of p_6 a sequence is realizable remains open.

Our experiments aimed to better understand and control the combinatorial information encoded by the p-vectors of simple 3-polytopes. To investigate such combinatorial properties, we utilized a dataset of 496 instances. 437 of these simple 3-polytopes originated from an online database of planar, 3-connected, 3-regular graphs (K. Coolsaet & Goedgebeur, 2023). The remaining 59 instances are randomly generated such graphs, wherein new random instances were accepted into the dataset only if they stretched the confines of the of the combinatorial data. That is to say, new instances were welcomed into the dataset of simple 3-polytopes if they occurred outside of the convex hull of the existing data (in regards to the combinatorial features of the p-vector and f-vector). This resulted in a non-redundant dataset, having the benefits of containing 'special' instances from theory (coming from (K. Coolsaet & Goedgebeur, 2023)) as well as random instances which expanded the boundaries of the data.

On this dataset of simple 3-polytopes, we examine the **IRIS** scores of four known inequalities which control the f and p vectors of a polytope (our features of choice). For each inequality, we examine its **IRIS** score with a constant weighting of $\boldsymbol{\lambda} = [1, 1, 1]^\top$ and provide a visual plot of the bound. Two notable lower bounds on p_6 include the following theorems.

Theorem 5.1. *(Barnette, 1969) Let P be a simple 3-polytope with $p(P) = (p_3, ..., p_m)$. If $\sum_{k\geq 7} p_k \geq 3$, then*

$$p_6 \geq 2 + \frac{p_3}{2} - \frac{p_5}{2} - \sum_{k\geq 7} p_k. \tag{2}$$

Theorem 5.2. *(Jucovič, 1971) Let P be a simple 3-polytope with $p(P) = (p_3, ..., p_m)$. If $f_2 \geq 7$, then*

$$p_6 \geq 4 - \frac{2}{3}p_4 - p_5 + \frac{1}{3}\sum_{k\geq 7}\left(\left\lfloor\frac{k+1}{2}\right\rfloor - 6\right)p_k. \tag{3}$$

We also investigate the classical Lower Bound Theorem by Barnette, which when formulated for simple 3-polytopes states $f_0 \geq 2f_2 - 4$ (Brøndsted, 1983). Finally, we examine the **IRIS** score of the well-known Hirsch Conjecture on our dataset of simple 3-polytopes. In this setting, the Hirsch Conjecture is known to be a theorem (see (Ziegler, 1995)) which states that $\mathrm{diam}(P) \leq f_2 - 3$ for 3-dimensional polytopes P and their combinatorial diameter.

Table 1. **IRIS** Scores of known inequalities

Statement h	$T(h;X)$	$R(h;X)$	$L(h;X)$
Barnette's Theorem	0.01754	1.0	0.14447
Juvovič's Theorem	0.01626	0.5	0.050759
Lower Bound Theorem If P is simple, then $f_0 \geq 2f_2 - 4$	0.99798	0.00404	1.0
Hirsch Conjecture If P is a 3-polytope, then $\mathrm{diam}(P) \leq f_2 - 3$	0.00806	0.5	0.007017

Barnette's Theorem In our dataset of 496 simple 3-polytopes, 171 instances meet the criteria of Barnette's Theorem 5.1. This is, in part, due to the dataset containing many m-gonal prisms and fullerenes.

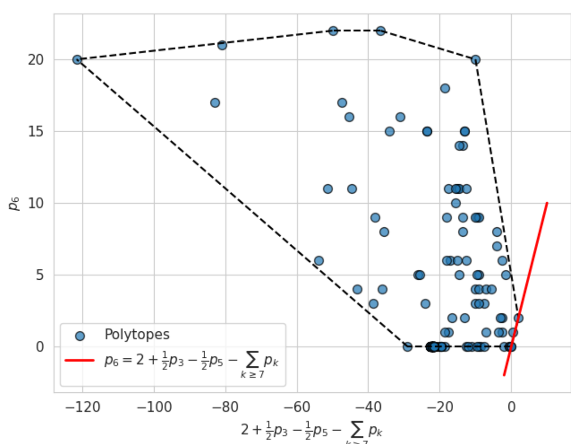

Figure 1. Barnette's lower bound on p_6. The black dashed line represents the convex hull of the plotted data, and the red line represents $p_6 = 2 + \frac{1}{2}p_3 - \frac{1}{2}p_5 - \sum_{k\geq 7} p_k$, with $\boldsymbol{\lambda}^\top \phi(h;X) \approx 1.16202$.

Barnette's Equation 2 lacks sharpness, and this small touch set both spans a low-dimensional affine space and encompasses a small geometric area when compared to the overall diameter of the data.

Jucovič's Theorem From our dataset of simple 3-polytopes, 492 instances satisfy conditions of Theorem 5.2.

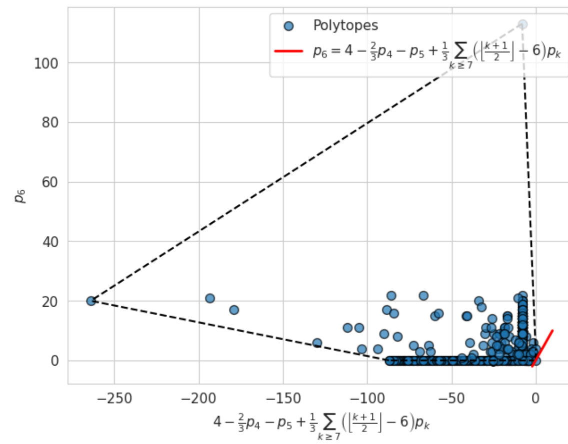

Figure 2. Jucovič's lower bound on p_6. The black dashed line represents the convex hull of the plotted data, and the red line represents $p_6 = 4 - \frac{2}{3}p_4 - p_5 + \frac{1}{3}\sum_{k\geq 7}\left(\lfloor\frac{k+1}{2}\rfloor - 6\right)p_k$, with $\boldsymbol{\lambda}^\top \phi(h;X) \approx 0.567019$.

Lower Bound Theorem The Lower Bound Theorem states that if P is a simple 3-polytope, then $f_0 \geq 2f_2 - 4$. That is to say, the number of vertices of P is bounded below in terms of the number of facets of P.

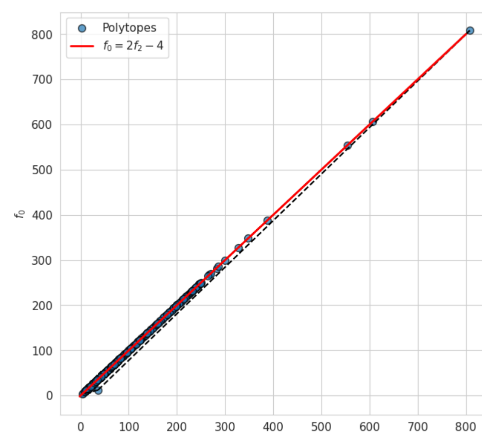

Figure 3. The Lower Bound Theorem on f_0. The black dashed line represents the convex hull of the plotted data, and the red line represents $f_0 = 2f_2 - 4$, with $\boldsymbol{\lambda}^\top \phi(h;X) \approx 2.002$.

The Lower Bound Theorem excels under the Normalized Touch and Normalized Diameter scores. This is due to sharpness being common and widespread; the shape of the convex hull of the dataset is largely encapsulated by this hyperplane.

Hirsch Conjecture The Hirsch Conjecture (again, a theorem in dimension 3) states that $\text{diam}(P) \leq f_2 - d$, wherein d is the dimension of polytope P. There are no assumptions on P, whereas the previous three theorems required P to be simple. That being said, simple polytopes are the 'worst case' scenario in terms of their diameter being large (Ziegler, 1995).

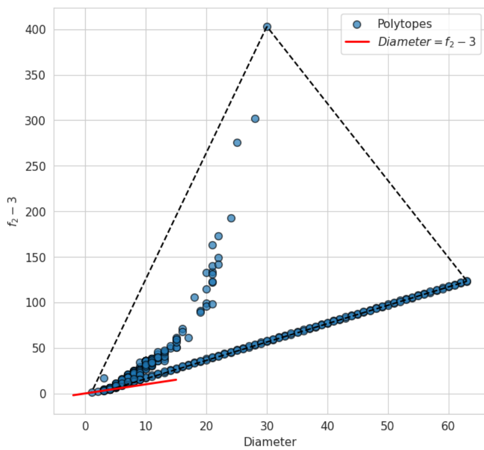

Figure 4. The Hirsch Conjecture in dimension 3. The black dashed line represents the convex hull of the plotted data, and the red line represents $\text{diam}(P) = f_2 - 3$, with $\boldsymbol{\lambda}^\top \phi(h; X) \approx 0.51508$.

Conjectured Bound We used **IRIS** to rank computer-developed conjectures in the spirit of Theorem 5.1. Our dataset of simple 3-polytopes satisfying the conditions of Barnette's theorem with invariants p_3, p_5, p_6 and $\sum_{k\geq 7} p_k$ was utilized. The convex hull of the dataset was taken, and **IRIS** was run on the identified face-defining hyperplanes with $\boldsymbol{\lambda} = [1, 1, 1]^\top$, $\boldsymbol{\beta} = [1, 1, 1, 1]^\top$, and $R = 100$. Among the resulting conjectures, we observed:

$$p_6 \geq \frac{39}{20} + \frac{p_3}{2} - \frac{p_5}{4} - \sum_{k\geq 7} p_k.$$

Table 2. **IRIS** Scores of Conjecture 6.1

$T(h; X)$	$R(h; X)$	$L(h; X)$	**Score**
0.029239	0.8	0.1444768	0.98123

The bound provided in Conjecture 6.1 is a stronger bound than Equation 2 when $p_5 > 0$.

5.2. Zero Forcing, Independence, and Domination in Subcubic Graphs

We applied NUEVAMIRADA to the family of connected graphs with maximum degree at most three, focusing on three foundational invariants: the *zero forcing number* $Z(G)$, the *domination number* $\gamma(G)$, and the *independence number* $\alpha(G)$. These parameters have served as benchmarks in prior work, including several conjectures produced by the original *TxGraffiti* system.

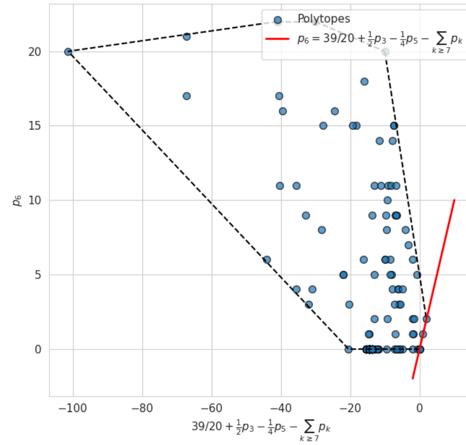

Figure 5. The conjectured lowered bound on p_6. The black dashed line represents the convex hull of the plotted data, and the red line represents $p_6 = \frac{39}{20} + \frac{1}{2}p_3 - \frac{1}{4}p_5 - \sum_{k\geq 7} p_k$, with $\boldsymbol{\lambda}^\top \phi(h; X) \approx 0.98123$.

To emphasize generality across Boolean hypotheses, we applied NUEVAMIRADA separately to multiple subfamilies of subcubic graphs, storing all valid conjectures. We then filtered and ranked them using a three-stage process: first, by retaining those with **IRIS** score $\boldsymbol{\lambda}^\top \phi(h; X) \geq 0.8$, second, by sorting with this score, and third, by applying the Morgan heuristic (Davila, 2024d) to keep only the most general version of each inequality across hypotheses; a system together called NUEVAMIRADA+.

When prompted to explore relationships between the zero forcing number and independence number, NUEVAMIRADA generated the inequality

$$Z(G) \leq \alpha(G) + 1,$$

for all connected subcubic graphs that are not complete graphs as its top-ranked conjecture (structural score $\boldsymbol{\lambda}^\top \phi(h; X) \approx 1.259$). This inequality was originally conjectured by *TxGraffiti* and is now widely believed to be true. It has since been proven for claw-free cubic graphs (Davila, 2019; Davila & Henning, 2020), and for almost all cubic graphs in (Schuerger et al., 2024).

In a separate experiment, we directed the system to examine potential inequalities involving the domination number and zero forcing number. Among the resulting statements was the inequality

$$\frac{1}{2}Z(G) \leq \gamma(G),$$

valid for all connected subcubic graphs not isomorphic to K_n (structural score $\boldsymbol{\lambda}^\top \phi(h; X) \approx 1.114$). When rearranged, this yields an upper bound on the zero forcing number for subcubic graphs: $Z(G) \leq 2\gamma(G)$. This statement generalizes a known theorem for connected cubic graphs not isomorphic to K_4—a result that originated as a conjecture from an early version of *TxGraffiti* and was later proven in (Davila, 2019; Davila & Henning, 2020).

Additionally, the system proposed the inequality

$$Z(G) \leq \gamma(G) + 2,$$

for all connected claw-free cubic graphs and all connected diamond-free cubic graphs, with both versions achieving a structural score of approximately $\boldsymbol{\lambda}^\top \phi(h; X) \approx 0.8439$. In both settings, this inequality had previously been conjectured by the original *TxGraffiti* system. The claw-free case has since been proven and fully characterized in (Davila, 2024c).

Among the inequalities surfaced by NuevaMirada, one notable conjecture appeared for connected cubic graphs

$$\frac{7}{12}Z(G) \leq \gamma(G) + \frac{49}{60}.$$

When rearranged, this yields an upper bound on the zero forcing number:

$$Z(G) \leq \frac{12}{7}\gamma(G) + \frac{7}{5},$$

a bound that was not sharp - a consequence of rounding to ratios in the floating point coefficients obtained via linear optimization over the convex hull of the invariant dataset. When looking at the true values, we see

$$0.5145 \cdot Z(G) - 0.8575 \cdot \gamma(G) \leq 0.686.$$

These decimal coefficients suggested the presence of an underlying rational structure. By observing that

$$\frac{0.8575}{0.5145} \approx \frac{5}{3}, \quad \text{and} \quad \frac{0.686}{0.5145} \approx \frac{4}{3}.$$

Thus, by again solving for $Z(G)$ with these approximations, we obtain

$$Z(G) \leq \frac{5}{3}\gamma(G) + \frac{4}{3},$$

as a possible upper bound on $Z(G)$ for connected cubic graphs. This refined conjecture is supported by **IRIS** scores and is presented in Figure 6.

This bound not only fits the data more tightly (sharp on the complete graph K_4) but also exhibits greater simplicity and interpretability—qualities often associated with meaningful mathematical structure. A visual representation of this inequality over the dataset of connected subcubic graphs is shown in Figure 6.

Table 3 summarizes the main inequalities discovered in these experiments.

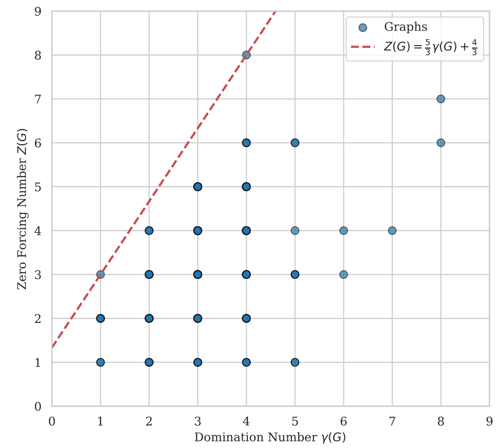

Figure 6. Upper bound on the zero forcing number in connected subcubic graphs: $Z(G) \leq \frac{5}{3}\gamma(G) + \frac{4}{3}$.

Inequality	Graph Family	Status
$Z \leq \alpha + 1$	$\Delta \leq 3,\ G \ncong K_n$	Known Conjecture
$Z \leq 2\gamma$	$\Delta \leq 3,\ G \ncong K_n$	New Conjecture
$Z \leq \gamma + 2$	Claw-free, cubic	Known Theorem
$Z \leq \gamma + 2$	Diamond-free, cubic	Known Conjecture
$Z \leq \frac{5}{3}\gamma + \frac{4}{3}$	$\Delta \leq 3$	New Conjecture

Table 3. Inequalities discovered by NuevaMirada in subcubic graphs involving zero forcing, domination, and independence.

6. Conclusion

IRIS provides a new framework for rapidly identifying mathematically meaningful conjectures. By combining geometric intuition with heuristic evaluation, the system is capable of rediscovering known results, refining established bounds, and surfacing novel inequalities—even when integrated with non-autonomous or hybrid workflows. This suggests a broader role for **IRIS** as a foundation for what might be called *data science for mathematical discovery*.

We conclude by formally stating several open conjectures proposed by NuevaMirada utilizing the **IRIS** scoring system during our study.

Conjecture 6.1 (*NuevaMirada*)**.** *If P is a simple 3-polytope with face vector $(p_3, p_4, \ldots, p_m)$, such that $\sum_{k\geq 7} p_k \geq 3$, then*

$$p_6 \geq \frac{39}{20} + \frac{p_3}{2} - \frac{p_5}{4} - \sum_{k\geq 7} p_k.$$

Conjecture 6.2 (Human and *NuevaMirada+*)**.** *If $G \neq K_4$ is a connected graph with maximum degree $\Delta \leq 3$, then*

$$Z(G) \leq \frac{5}{4}\gamma(G) + \frac{4}{3}.$$

Conjecture 6.3 (*NuevaMirada+*)**.** *If G is a connected graph* *with maximum degree* $\Delta \leq 3$*, then*

$$Z(G) \leq 2\gamma(G).$$

Impact Statement

This paper presents work whose goal is to advance the field of machine-assisted mathematical discovery. While there may be long-term societal implications—particularly regarding the evolving role of the mathematician—we do not believe any specific concerns need to be highlighted at this time.

Acknowledgements:

We are grateful for the financial support received from NSF grant 2434665. Ethan Fang is partially supported by NSF grants DMS-2346292 and DMS-2434666. Junwei Lu is partially supported by NSF DMS-2434664. RelationalAI provided funding for presenting this work.

References

Barnette, D. On p-vectors of 3-polytopes. *J. Combin. Theory*, 1969.

Brimkov, B., Davila, R., Schuerger, H., and Young, M. On a conjecture of TxGraffiti: Relating zero forcing and vertex covers in graphs. *Discrete Appl. Math.*, 359:290–302, 2024.

Brøndsted, A. *An Introduction to Convex Polytopes*. Springer Science+Business Media New York, 1983.

Caro, Y., Davila, R., and Pepper, R. New results relating matching and independence. *Discuss. Math. Graph Theory*, 42(3):921–935, 2022.

Chung, F. The average distance and the independence number. *J. Graph Theory*, 12(2):229–235, 1988.

Davies, A., Juhász, A., Lackenby, M., Tomašev, N., Blundell, C., Veličković, P., Buesing, L., et al. Advancing mathematics by guiding human intuition with ai. *Nature*, 600(7887):70–74, 2021. doi: 10.1038/s41586-021-04086-x.

Davila, R. Total and zero forcing in graphs, 2019. PhD Thesis.

Davila, R. Automated conjecturing in mathematics with TxGraffiti. *arXiv preprint arXiv:2409.19379v1*, 2024a. URL https://arxiv.org/abs/2409.19379.

Davila, R. TxGraffiti online: An interactive platform for automated conjecture generation, 2024b. URL https://txgraffiti.streamlit.app. Accessed April 2024.

Davila, R. Another conjecture of TxGraffiti concerning zero forcing and domination in graphs, 2024c. URL https://arxiv.org/abs/2406.19231.

Davila, R. The Optimist: Towards fully automated graph theory research, 2024d. URL https://arxiv.org/abs/2411.09158.

Davila, R. and Henning, M. Total forcing versus total domination in cubic graphs. *Appl. Math. Comput.*, 354:385–395, 2019.

Davila, R. and Henning, M. Zero forcing in claw-free cubic graphs. *Bull. Malays. Math. Sci. Soc.*, 43:673–688, 2020.

Davila, R. and Henning, M. Zero forcing versus domination in cubic graphs. *J. Comb. Optim.*, 41:553–577, 2021.

DeLaViña, E. Some history of the development of Graffiti. In *Graphs and Discovery, DIMACS Ser. Discret. Math. Theor. Comput. Sci.*, volume 69, pp. 81–118, Providence, RI, 2005. Amer. Math. Soc.

DeLaViña, E., Pepper, R., and Waller, B. Independence, radius, and hamiltonian paths: New conjectures from Graffiti.pc. *Preprint*, 2024.

Devillez, G., Hauweele, P., and Mélot, H. PHOEG Helps to Obtain Extremal Graphs. In *Operations Research Proceedings 2018: Selected Papers of the Annual International Conference of the German Operations Research Society*, pp. 251–257, Cham, Switzerland, 2019. Springer Nature.

Fajtlowicz, S. On conjectures of Graffiti, II. *Congr. Numer.*, 60, 1987.

Fajtlowicz, S. A characterization of radius-critical graphs. *J. Graph Theory*, 12, 1988.

Fajtlowicz, S. On conjectures of Graffiti, IV. *Congr. Numer.*, pp. 231–240, 1990.

Fajtlowicz, S. On representation and characterization of buckminsterfullerene c60. In *Graphs and Discovery DIMACS: Ser. Discrete Math. Theor. Comput. Sci.*, volume 69, pp. 127–135. 2005.

Fajtlowicz, S. and Larson, C. Graph-theoretical independence as a predictor of fullerene stability. *Chem. Phys. Lett.*, 377:485–490, 2003.

Fajtlowicz, S., John, P., and Sach, H. On maximum matchings and eigenvalues of benzenoid graphs. *Croat. Chem. Acta*, 78:195–201, 2005.

Fowler, P. Fullerene graphs with more negative than positive eigenvalues; the exceptions that prove the rule of electron deficiency. *J. Chem. Soc. Faraday*, 93:1–3, 1997.

Fowler, P., Hansen, P., Rogers, K. M., and Fajtlowicz, S. C60br24 as a chemical illustration of graph theoretical independence. *J. Chem. Soc. Perkin Trans.*, 2, 1998.

Fowler, P., Rodgers, K., Fajtlowicz, S., Hansen, P., and Caporossi, G. Facts and conjectures about fullerene graphs. leapfrog, cylinder and fullerene graphs. In *Proc. Euroconf. ALCOMA*, 1999.

Grünbaum, B. *Convex Polytopes*. Springer-Verlag New York, Inc., 2003.

Jucovič, E. On the number of hexagons in a map. *J. Combin. Theory*, 1971.

K. Coolsaet, S. D. and Goedgebeur, J. House of graphs 2.0: A database of interesting graphs and more, 2023. URL https://houseofgraphs.org.

Larson, C. E. and Cleemput, N. V. Automated conjecturing I: Fajtlowicz's dalmatian heuristic revisited. *Artif. Intell.*, 231:17–38, 2016.

Mélot, H. Facets of the graph invariant polytope: Application to the automated discovery of conjectures. *Discrete Math.*, 273:119–137, 2003.

Mélot, H. Facet defining inequalities among graph invariants: The system GraPHedron. *Discrete Appl. Math.*, 156: 1875–1891, 2008.

Pepper, R. On new didactics of mathematics: Learning graph theory via Graffiti. In *Graphs and Discovery*, volume 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pp. 341–349, Providence, RI, 2001. American Mathematical Society.

Schuerger, H., Warnberg, N., and Young, M. Zero forcing and vertex independence number on cubic and subcubic graphs. *arXiv preprint arXiv:2410.21724*, 2024. URL https://arxiv.org/abs/2410.21724.

Whitney, H. 2-isomorphic graphs. *American Journal of Mathematics*, 55(1):245–254, 1933. ISSN 00029327, 10806377. URL http://www.jstor.org/stable/2371127.

Ziegler, G. M. *Lectures on Polytopes*. Springer-Verlag New York, Inc., 1995.