GAMBIT: Generating Automated Mathematical Bounds, Inequalities, and Theorems

Randy Davila
Research and Development
RelationalAI
Berkeley CA, United States

Department of Computational Applied
Mathematics & Operations Research
Rice University
Houston TX, United States
rrd6@rice.edu

Abstract

Conjecture and proof are the twin engines of mathematics. Despite rapid progress in automated proof, conjecture remains underexplored. Early efforts, beginning with *Graffiti* in the 1980s and followed by only a handful of systems, showed that computer-generated conjectures could enter the literature, but all depended on user-specified targets and thus assumed domain expertise. To complement modern large language models and automated proof tools, we need systems that can autonomously decide what quantities are worth relating. We present GAMBIT (*Generating Automated Mathematical Bounds, Inequalities, and Theorems*), which identifies promising quantities and hypotheses in structured datasets, then applies optimization-based searches and heuristic filters to propose candidate relations. Case studies in graphs and Calabi–Yau varieties show that GAMBIT both rediscovers known results and suggests novel ones, illustrating AI-driven conjecturing as a timely complement to advances in theorem proving and counterexample search.

1 Introduction

Mathematics advances through a dialogue between *conjecture* and *proof*. Conjectures articulate hidden structure and guide exploration, while proofs transform conjectures into knowledge. In recent years, AI has made remarkable progress on the proving side, through automated theorem provers [Harrison, 2009, de Moura et al., 2015], interactive proof assistants with rapidly growing libraries [The mathlib Community, 2019], and large language models that extend formal reasoning [Polu and Sutskever, 2020, Yang et al., 2023a]. In contrast, the creative step of *conjecturing*, long central to mathematical practice, remains comparatively neglected by the AI community. This imbalance motivates a renewed effort to explore how computers might generate conjectures.

Early systems demonstrated that computers could propose conjectures worthy of mathematical study. This line of work began with *Graffiti* in the 1980s [Fajtlowicz, 1988a, 1987, 1988b, 1990, 1995], and was later continued by systems such as *Graffiti.pc* [DeLaViña, 2005], *Conjecturing* [Larson and Cleemput, 2016], *AutoGraphiX* [Caporossi and Hansen, 2000], *GraPHedron* [Mélot, 2008], *PHOEG* [Devillez et al., 2019], and *TxGraffiti* [Davila, 2024]. These programs generated inequalities that entered the mathematical literature. Because they required users to nominate a target parameter, these systems excelled when applied to familiar areas of mathematics. They offered less guidance, however, in domains where the relevant invariants were not yet clear—an aspect that motivates our approach.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI.

We introduce GAMBIT (*Generating Automated Mathematical Bounds, Inequalities, and Theorems*), a new class of AI-driven conjecturing system that removes this restriction. Rather than fixing a single target, GAMBIT systematically considers all available invariants, ranks promising quantities and structural conditions, and proposes candidate equalities and inequalities through optimization-based searches. These candidates are filtered by heuristic tests of truth, sharpness, and significance, yielding statements that resemble the kinds of conjectures mathematicians formulate. In this way, GAMBIT enables *domain-general conjecturing* from structured datasets, including graphs, integers, matrices, and Calabi–Yau varieties, without requiring prior domain knowledge.

The name GAMBIT is deliberate: in chess, a gambit sacrifices a pawn to open new possibilities. Here, the system sacrifices foreknowledge of what to target, proposing conjectures across all invariants without fixing a goal in advance. This embodies both the risk and the promise of discovery. This capability complements recent advances in AI for proof, including Minerva [Lewkowycz et al., 2022], Llemma [Azerbayev et al., 2024], LeanDojo's retrieval-augmented proving [Yang et al., 2023b], and AlphaGeometry for geometry [Trinh et al., 2024]. Whereas these systems extend the proving side of mathematics, GAMBIT addresses the complementary task: producing plausible statements and structural conditions that can guide both human reasoning and automated proof search.

The remainder of the paper presents the design of GAMBIT, demonstrates its results on graphs and Calabi–Yau datasets, and discusses implications for AI in mathematics. More broadly, we suggest that automated conjecturing should be viewed as a central challenge and a necessary complement to AI advances in theorem proving and counterexample search.

2 GAMBIT Design

Mathematicians often form conjectures by noticing regularities: one quantity consistently bounds another, two invariants coincide on special subfamilies, or a relation sharpens under a structural condition. GAMBIT automates this kind of pattern-seeking on structured datasets. Given a table of objects annotated with numeric invariants and Boolean predicates, GAMBIT searches for regularities expressible as precise relations and then filters them for plausibility. The aim is discovery rather than prediction: crisp, interpretable statements that resemble the conjectures mathematicians formulate.

To make this concrete, GAMBIT represents each conjecture as an *invariant–hypothesis–conclusion triple* of the form

$$h \;\Rightarrow\; f(\text{invariants}) \;\bowtie\; g(\text{invariants}),$$

where h is a Boolean condition (possibly trivial), $\bowtie \in \{\leq, \geq, =\}$, and f, g are affine expressions in the available invariants. In the dataset, rows correspond to objects, numeric columns to invariants, and Boolean columns to candidate hypotheses for conditional statements. Let $\mathcal{D}$ denote the set of rows in the dataset (objects). For any Boolean predicate h, the induced slice is

$$\mathcal{D}_h = \{\, i \in \mathcal{D} : h(i) = \text{true} \,\}, \qquad c(h) \doteq |\mathcal{D}_h|/|\mathcal{D}|,$$

where $c(h)$ denotes the coverage of h, i.e. the fraction of rows it selects. Conjecturing thus reduces to a systematic search for triples of this form that are satisfied for all objects in $\mathcal{D}_h$.

GAMBIT begins by ranking pairs of invariants and identifying slices $\mathcal{D}_h$ under which their relationship becomes more strongly correlated or tightly bounded. For any subset $S \subseteq \mathcal{D}$ (typically $S = \mathcal{D}$ or $S = \mathcal{D}_h$), GAMBIT evaluates the degree of alignment of two invariants by computing two complementary signals. The first is an absolute correlation measure

$$\operatorname{corr}_S(u, v) = \left| \frac{\sum_{i \in S}(u_i - \overline{u}_S)(v_i - \overline{v}_S)}{\sqrt{\sum_{i \in S}(u_i - \overline{u}_S)^2}\,\sqrt{\sum_{i \in S}(v_i - \overline{v}_S)^2}} \right|,$$

computed using only finite entries. The second is a ratio-tightness score,

$$\operatorname{tight}_S(u, v) \;=\; \begin{cases} \dfrac{\min_{i \in I_S} |u_i/v_i|}{\max_{i \in I_S} |u_i/v_i|}, & \text{if } \operatorname{sign}(u_i/v_i) \text{ is constant on } I_S, \\ 0, & \text{otherwise,} \end{cases}$$

where $I_S = \{\, i \in S : |v_i| \geq \varepsilon,\; (u_i, v_i) \text{ finite} \,\}$ and $\varepsilon = 10^{-9}$ avoids division by zero. By construction, $\operatorname{tight}_S \in [0, 1]$, attaining 1 exactly when the ratio u/v is constant on S.

This affinity score highlights slices where two invariants exhibit both strong correlation and near-constant ratios, conditions that often signal underlying mathematical structure. GAMBIT defines affinity by the weighted geometric mean

$$\mathrm{affinity}_S(u,v) = \mathrm{corr}_S(u,v)^{\alpha}\,\mathrm{tight}_S(u,v)^{\,1-\alpha},$$

with $\alpha \in [0,1]$ (we use $\alpha = 0.7$ in all experiments). Not every slice is useful: some may be too small, too broad, or fail to reveal a stronger structure than the full dataset. A slice induced by h is therefore considered *informative* for (u,v) only if (i) its coverage lies in $[\underline{c}, \overline{c}]$ with $\underline{c} = 0.05$, $\overline{c} = 0.95$, and at least $N_{\min} = 20$ rows, and (ii) its affinity improves upon the global dataset by at least $\eta = 0.10$. Informative pairs (u,v) together with their slices h define ranked *plans* that seed the subsequent generation of conjectures via linear programming, ratio searches, or convex-hull approximations.

For each informative slice $\mathcal{D}_h$, GAMBIT proposes explicit inequalities and equalities by solving a linear program restricted to that slice. Fix a target invariant y and a vector of partner invariants $x \in \mathbb{R}^k$. On $\mathcal{D}_h$, GAMBIT solves an ℓ_∞ slack–minimization problem to search for affine bounds. For example, to obtain an *upper bound* $y \le a^\top x + b$, it solves

$$\begin{aligned}
\min_{a,b,M} \quad & M \\
\text{s.t.} \quad & a^\top x_i + b \;\ge\; y_i \quad \forall i \in \mathcal{D}_h, \\
& (a^\top x_i + b) - y_i \;\le\; M \quad \forall i \in \mathcal{D}_h, \\
& M \;\ge\; 0,
\end{aligned}$$

whose solution yields the conditional inequality $h \Rightarrow y \le a^\top x + b$. The formulation for a *lower bound* $y \ge a^\top x + b$ is symmetric, reversing the inequalities. In both cases, the slack variable M bounds the maximum violation on the slice, and minimizing M returns the tightest feasible affine inequality. Equalities arise when both directions hold simultaneously.

Even after slice selection, many raw inequalities are redundant or can be generalized to larger families of objects. To address this, GAMBIT applies a postprocessing stage. First, inequalities are normalized for readability: coefficients are simplified, negatives are moved consistently to the left-hand side, and equalities are inferred when both bounds coincide. The normalized set is then filtered: redundant statements are discarded, overly specific hypotheses are generalized when possible, and significance is enforced by requiring that each surviving conjecture provide a strictly better bound than all previously accepted ones for at least one object. Finally, conjectures are ranked by *touch count*—the number of objects on which equality holds—so that sharp and frequently realized statements appear first.

3 Case Studies: Graphs and Calabi–Yau Manifolds

To illustrate the range of GAMBIT, we present results in two contrasting domains: graph theory and Calabi–Yau geometry. In both cases, the input was simply a table of objects annotated with numeric and Boolean invariants, with no user-specified targets or hypotheses. All conjectures emerged automatically from the pipeline. A companion Colab notebook provides reproducibility.[1].

A central methodological distinction between automated conjecturing and standard machine learning lies in data usage. Whereas machine learning pipelines typically partition data into training, validation, and test sets, automated conjecturing systems operate on the complete dataset. This reflects a mathematical rather than statistical rationale: like a human mathematician, an automated conjecturer benefits from considering all known examples when seeking new relationships. Omitting portions of the data risks excluding extremal or atypical instances that may reveal structural boundaries or point toward counterexamples. Although conjectures formed in this manner may not generalize beyond the current dataset, they often expose meaningful patterns and motivate further theoretical inquiry.

Graph theory conjectures

Experiments used an open-source PyPI package providing a dataset of 340 graphs annotated with over twenty numerical invariants and a dozen Boolean predicates. The invariants span structural,

[1] `https://colab.research.google.com/drive/1nGGso4B7_4GtUSY4sETBIgpg0d74u0X5?usp=sharing`

domination, spectral, and forcing parameters, while the predicates capture properties such as connectedness, bipartiteness, chordality, planarity, and subcubic degree. Although all features were made available, the system automatically deprioritized those that did not yield informative slices. Conjectures are evaluated by *touch count*, the number of graphs on which equality holds, with higher counts indicating sharper and more frequently realized statements.

Category	**Representative conjectures (touch count)**
Coloring/cliques	$\chi \geq \omega$ (296); chordal $\Rightarrow \chi = \omega$ (296)
Domination	$\gamma \leq i$ (321); $\gamma \geq sl$ (224); $\gamma \leq \gamma_t$ (114); $\gamma_t \leq 2\, sl$ (93); $\gamma \geq \frac{1}{2} i + \frac{1}{4}\gamma_t$ (73); planar $\Rightarrow sl \geq \frac{1}{6}\gamma + \frac{1}{6} i + \frac{2}{3}$ (54); $i \leq \frac{7}{8}\gamma + \frac{5}{8} sl - \frac{1}{2}$ (49)
Matching/vertex cover	$\mu \leq \nu$ (175); bipartite $\Rightarrow \mu = \nu$ (175); $\mu_{\text{min -max}} \geq \frac{1}{2}\nu$ (67); $\mu_{\text{min -max}} \leq -\frac{1}{3}\nu + \frac{4}{3} + \mu$ (17); $m \geq 2\nu - 1$ (24)
Ann./independence/residue	$a \geq \alpha$ (138); $\alpha \geq R$ (105); chordal $\Rightarrow R \geq \frac{1}{2}\alpha + \frac{1}{2}$ (99); $a \leq \frac{1}{2}\alpha + \frac{1}{2} n - \frac{1}{2}$ (45)
Distance	$\text{rad} \geq \frac{1}{2}\text{diam}$ (115); $\text{diam} \geq \frac{1}{3}\text{tri}$ (74); $\text{tri} \geq 4\,(\text{diam} - \text{rad})$ (16); chordal $\Rightarrow \text{rad} \leq \frac{1}{3}\text{diam} + \frac{1}{2} sl + \frac{1}{3}$ (45)
Harmonic index	$H \leq \frac{1}{2} n$ (134); subcubic $\Rightarrow m \leq 3H$ (121)

Table 1: Representative conjectures on graph invariants generated by GAMBIT. Numbers in parentheses denote *touch counts*, the number of graphs on which equality holds.

Table 1 shows representative conjectures spanning domination, coloring, matching, and distance theory. Several textbook theorems appeared prominently [West, 2001], including $\chi \geq \omega$, Kőnig's theorem ($\mu = \nu$ in bipartite graphs), and the radius–diameter inequality $\text{rad} \geq \frac{1}{2}\text{diam}$, as well as specialized and technical theorems such as $\gamma \geq sl$ [Haynes et al., 1998], $\text{diam} \geq \frac{1}{3}\text{tri}$ [Das, 2021], $a \geq \alpha$ [Pepper, 2009], and $\alpha \geq R$ [Favaron et al., 1991, Griggs and Kleitman, 1999, Triesch, 1996]. In addition, GAMBIT proposed candidates that appear to be new, such as $\gamma \geq \frac{1}{2} i + \frac{1}{4}\gamma_t$ and $\text{rad} \leq \frac{1}{3}\text{diam} + \frac{1}{2} sl + \frac{1}{3}$ in chordal graphs. In fact, $\gamma \geq \frac{1}{2} i + \frac{1}{4}\gamma_t$ holds in the *claw-free* graphs.

This mixture of rediscoveries, less familiar results, and novel conjectures illustrates how the system converts raw tabular data into statements that warrant mathematical scrutiny. At the same time, many invariants did not yield conjectures, which underscores that GAMBIT is selective rather than exhaustive.

Calabi–Yau conjectures

For Calabi–Yau hypersurfaces, we used an open-source PyPI package that derives features from `cy-tools` [Demirtas et al., 2022]. Each row corresponds to a generic anticanonical hypersurface in a toric variety X_Δ. The dataset provides algebraic invariants ($h^{1,1}, h^{2,1}, \chi$), graph statistics from intersection and Mori-ray angle graphs, matrix-derived features from intersection forms, and Boolean predicates such as connectivity and density > 0.5. In total, it includes 175 Calabi–Yau threefolds with 44 numerical and 19 Boolean features; see Appendix A for exact definitions.

Category	**Representative conjectures (touch count)**
Rediscoveries	$n_{\text{mori\_nodes}} = n_{\text{mori\_rays}}$ (175); $h^{1,1} = \frac{1}{2} n_{\text{int}} + \frac{1}{2} d_{\text{tip}}$ (175)
Structural	$\text{intersection\_rank} \leq \frac{1}{2} h^{1,1} + \frac{1}{2} d_{\text{tip}}$ (174); $\alpha_{\text{int}} \geq R_{\text{int}}$ (102); $\gamma_{\text{int}} \leq R_{\text{int}}$ (71); $\chi \geq -2h^{2,1} + 2 + 2\bar{d}_{\text{int}}$ (57); $h^{2,1} \geq -\frac{1}{2}\chi + 1 + \bar{d}_{\text{int}}$ (57)
Hybrids	$\text{int\_density} > 0.5 \Rightarrow \text{intersection\_rank} \geq \frac{7}{10} h^{1,1} - 4 + \frac{7}{10} d_{\text{tip}}$ (26); Mori forest $\Rightarrow \gamma_{\text{mori}} \leq \frac{1}{3} n_{\text{mori\_components}} + \frac{2}{3} + \frac{24}{17}\cos\theta_{\max}$ (4); $\text{intersection\_sum\_abs} \leq \frac{17}{7} + S_{\text{trace}} + \frac{3}{7} n_{\text{intersections}}$ (2)

Table 2: Representative conjectures from Calabi–Yau datasets. Numbers in parentheses denote *touch counts*, the number of instances on which equality holds.

Here, GAMBIT rediscovers exact identities that serve as consistency checks, surfaces structural inequalities linking algebraic and graph-theoretic quantities, and proposes hybrid conditional statements tied to density or spectral features. As in the graph case, several invariants did not produce conjectures, highlighting the system's selectivity.

4 Discussion

Our contribution is twofold. First, GAMBIT treats conjecturing as a task distinct from prediction and proof: the aim is to surface sharp, interpretable regularities rather than forecast outcomes or verify statements. Second, it shows domain-general scope: from structured datasets alone, the system both rediscovers textbook theorems and suggests relations that appear new. The next challenge is not only to expand the range of domains but to deepen the evaluation of conjectures: deciding which statements are genuinely structural, how significance should be measured, and how automated conjecturing might interact with proof systems. Addressing these questions will determine whether conjecturing can become a systematic component of mathematical discovery. For now, GAMBIT should be seen as a first move—an opening that sets the stage for further play.

A Definitions of Graph-Based Features for Calabi–Yau Datasets

In this section we describe the graphs and associated invariants used in the conjectures of Table 2.

A.1 Algebraic invariants

- $h^{1,1}$: number of Kähler parameters (divisor classes).
- $h^{2,1}$: number of complex structure deformations.
- $\chi = 2(h^{1,1} - h^{2,1})$: Euler characteristic of the threefold.
- d_{tip}: Kähler tip dimension, i.e., the dimension of the cone tip in the nef cone.

A.2 Graph Constructions

- **Intersection graph.** Let $\{D_i\}_{i=1}^{h^{1,1}}$ denote the divisor classes on the Calabi–Yau threefold X. The *intersection graph* G_{int} has vertex set $V(G_{\text{int}}) = \{D_i\}$, with an edge $\{D_i, D_j\} \in E(G_{\text{int}})$ whenever the intersection number

$$\kappa_{ijk} = \int_X D_i D_j D_k$$

  is nonzero for some k. Thus, divisors that intersect nontrivially are adjacent in G_{int}.
- **Mori graph.** Let $\{r_j\}$ denote the primitive rays generating the Mori cone of X. The *Mori graph* G_{mori} has vertex set $V(G_{\text{mori}}) = \{r_j\}$, with an edge $\{r_i, r_j\}$ present when the angle between rays r_i and r_j is acute. We record both the unweighted structure (edges only) and angle-based statistics, including

$$\cos\theta_{\max} = \max_{i \neq j} \cos\angle(r_i, r_j).$$

A.3 Graph Invariants

For a graph $G = (V, E)$ with $|V| = n$ and $|E| = m$, we use the following invariants:

- n: number of vertices.
- **Independence number** $\alpha(G)$: maximum size of a set of pairwise nonadjacent vertices.
- **Domination number** $\gamma(G)$: minimum size of a set $S \subseteq V$ such that every $v \in V \setminus S$ has a neighbor in S.
- **Residue** $R(G)$: the terminal value obtained by iteratively applying the Havel–Hakimi degree-reduction process.
- **Density** $\rho(G) = \frac{2m}{n(n-1)}$: proportion of possible edges present.

- **Components** $c(G)$: number of connected components.

Dataset notation:

- n_{int}: vertices in the intersection graph.
- $n_{\text{mori\_nodes}}$: vertices in the Mori graph.
- $n_{\text{mori\_rays}}$: Mori cone rays (equal to $n_{\text{mori\_nodes}}$).
- $n_{\text{mori\_components}}$: components of the Mori graph.
- α_{int}: independence number of the intersection graph.
- γ_{int}: domination number of the intersection graph.
- R_{int}: Havel-Hakimi residue of the intersection graph.
- γ_{mori}: domination number of the Mori graph.
- ρ_{int}: density of the intersection graph.
- $\cos\theta_{\text{max}}$: maximum cosine of a pairwise Mori ray angle.
- $\overline{d}_{\text{int}}$: average degree of the intersection graph.

A.4 Matrix-Derived Features

From the totally symmetric triple intersection tensor $\kappa_{ijk} = \int_X D_i D_j D_k$, we form contracted intersection matrices and extract:

- intersection_rank: rank of the intersection matrix.
- $\text{intersection\_sum\_abs}$: sum of absolute values of matrix entries.
- S_{trace}: trace of the squared intersection matrix.
- $n_{\text{intersections}}$: number of nonzero entries in the intersection matrix.

A.5 Boolean Predicates

- **Mori forest**: the Mori graph G_{mori} is acyclic.
- **Intersection density** > 0.5: $\rho(G_{\text{int}}) > 0.5$.
- **Connectivity**: the graph is connected.
- **Acyclicity**: the graph contains no cycles.

References

Zhangir Azerbayev, Hailey Schoelkopf, Keiran Paster, Marco Dos Santos, Stephen McAleer, Albert Q. Jiang, Jia Deng, Stella Biderman, and Sean Welleck. Llemma: An open language model for mathematics. In *International Conference on Learning Representations (ICLR)*, 2024. URL https://arxiv.org/abs/2310.10631. arXiv preprint arXiv:2310.10631.

Gilles Caporossi and Pierre Hansen. Variable neighborhood search for extremal graphs. I. the AutoGraphiX system. *Discrete Mathematics*, 212(1–2):29–44, 2000.

Angsuman Das. Triameter of graphs. *"Discuss. Math. Graph Theory*, 41(3):601–616, 2021. doi: 10.7151/dmgt.2212. URL https://doi.org/10.7151/dmgt.2212.

Randy Davila. Automated conjecturing in mathematics with TxGraffiti. *arXiv preprint arXiv:2409.19379*, 2024. doi: 10.48550/arXiv.2409.19379. URL https://arxiv.org/abs/2409.19379.

Leonardo de Moura, Soonho Kong, Jeremy Avigad, Floris van Doorn, and Jakob von Raumer. The lean theorem prover (system description). In *Automated Deduction – CADE 25*, volume 9195 of *Lecture Notes in Computer Science*, pages 378–388. Springer, 2015. doi: 10.1007/978-3-319-21401-6_26.

Ermelinda DeLaViña. Graffiti.pc: A variant of Graffiti. In *Graphs and Discovery*, volume 69 of *DIMACS Series in Discrete Mathematics and Theoretical Computer Science*, pages 71–79. American Mathematical Society, Providence, RI, 2005.

Mehmet Demirtas, Andres Rios-Tascon, and Liam McAllister. Cytools: A software package for analyzing calabi-yau manifolds. 2022. doi: 10.48550/arXiv.2211.03823. URL https://arxiv.org/abs/2211.03823.

Gauvain Devillez, Pierre Hauweele, and Hadrien Mélot. PHOEG helps to obtain extremal graphs. In Bernard Fortz and Martine Labbé, editors, *Operations Research Proceedings 2018: Selected Papers of the Annual International Conference of the German OR Society (GOR)*, pages 251–258. Springer, Cham, 2019.

Siemion Fajtlowicz. On conjectures of Graffiti II. *Congr. Numer.*, 60:187–197, 1987.

Siemion Fajtlowicz. On conjectures of Graffiti. *Discrete Math.*, 72(1–3):113–118, 1988a. doi: 10.1016/0012-365X(88)90199-9.

Siemion Fajtlowicz. On conjectures of Graffiti III. *Congr. Numer.*, 66:23–32, 1988b.

Siemion Fajtlowicz. On conjectures of Graffiti IV. *Congr. Numer.*, 70:231–240, 1990.

Siemion Fajtlowicz. On conjectures of Graffiti V. In Yousef Alavi and Allen Schwenk, editors, *Graph Theory, Combinatorics, and Algorithms (Proceedings of the 7th Quadrennial International Conference on the Theory and Applications of Graphs, Vol. 1)*, pages 367–376, New York, 1995. Wiley.

O. Favaron, M. Mahéo, and J.-F. Saclé. On the residue of a graph. *J. Graph Theory*, 15:39–64, 1991.

J. R. Griggs and D. J. Kleitman. Independence and the Havel–Hakimi residue. *Discrete Math.*, 127: 241–249, 1999.

John Harrison. Hol light: An overview. In *Theorem Proving in Higher Order Logics (TPHOLs)*, volume 5674 of *Lecture Notes in Computer Science*, pages 60–66. Springer, 2009. doi: 10.1007/978-3-642-03359-9_4.

Teresa W. Haynes, Stephen T. Hedetniemi, and Peter J. Slater. *Fundamentals of Domination in Graphs*. CRC Press, 1998.

Craig E. Larson and Nico Van Cleemput. Automated conjecturing I: Fajtlowicz's dalmatian heuristic revisited. *Artificial Intelligence*, 231:17–38, 2016.

Aitor Lewkowycz, Anders Andreassen, David Dohan, Ethan Dyer, Henryk Michalewski, Vinay Ramasesh, Ambrose Slone, Cem Anil, Imanol Schlag, et al. Solving quantitative reasoning problems with language models. In *NeurIPS*, 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/18abbeef8cfe9203fdf9053c9c4fe191-Paper-Conference.pdf.

Hadrien Mélot. Facet defining inequalities among graph invariants: The system GraPHedron. *Discrete Applied Mathematics*, 156(10):1875–1891, 2008.

Ryan Pepper. On the annihilation number of a graph. In *Proceedings of the 15th American Conference on Applied Mathematics (ACAM)*, volume 1, pages 217–220. WSEAS, 2009.

Stanislas Polu and Ilya Sutskever. Generative language modeling for automated theorem proving. *arXiv preprint arXiv:2009.03393*, 2020. URL https://arxiv.org/abs/2009.03393.

The mathlib Community. The Lean mathematical library. *arXiv preprint arXiv:1910.09336*, 2019. URL https://arxiv.org/abs/1910.09336.

Eberhard Triesch. Degree sequences of graphs and dominance order. *J. Graph Theory*, 22:89–93, 1996.

Trieu H. Trinh, Yuhuai Wu, Quoc V. Le, He He, and Thang Luong. Solving olympiad geometry without human demonstrations. *Nature*, 625(7950):476–482, 2024. doi: 10.1038/s41586-023-06747-5. URL https://www.nature.com/articles/s41586-023-06747-5.

Douglas B. West. *Introduction to Graph Theory*. Prentice Hall, 2nd edition, 2001.

Kaiyu Yang, Aidan M. Swope, Alex Gu, Rahul Chalamala, Peiyang Song, Shixing Yu, Saad Godil, Ryan Prenger, and Anima Anandkumar. Leandojo: Theorem proving with retrieval-augmented language models. *arXiv preprint arXiv:2306.15626*, 2023a. URL https://arxiv.org/abs/2306.15626.

Kaiyu Yang, Aidan M. Swope, Alex Gu, Rahul Chalamala, Peiyang Song, Shixing Yu, Saad Godil, Ryan Prenger, and Anima Anandkumar. Leandojo: Theorem proving with retrieval-augmented language models. In *NeurIPS Datasets and Benchmarks*, 2023b. URL https://papers.neurips.cc/paper_files/paper/2023/file/4441469427094f8873d0fecb0c4e1cee-Paper-Datasets_and_Benchmarks.pdf.