Lemmas and conjectures in spectral graph theory
from the automated analysis of the Duhamel–Cazenave corpus

Tristan Cazenave

October 31, 2026

Abstract

The corpus of Elie Duhamel and Tristan Cazenave [1] contains 305,755 automatically generated conjectures relating spectral and metric invariants of connected graphs. A pipeline (numerical refutation, proof by the SMT solver `z3` from a library of lemmas, discovery of new lemmas, formalisation in `Lean` 4/Mathlib) refuted 162,964 conjectures and proved 117,390 without any unproved hypothesis, of which 79,505 are verified end to end in `Lean` for every connected graph. This document collects, with proofs written for mathematicians, the most interesting lemmas that were discovered and proved, the ten most useful lemmas, the ten hardest theorems, simpler proofs of known results (Bollobás–Erdős, Yang–Lu), and the unrefuted conjectures that we find most interesting. All the proofs are gathered in an archive (Section 8).

1 Setting and notation

All graphs are simple, undirected and *connected*, with $n \geq 2$ vertices and m edges. We write d_v for the degree of v, Δ and δ for the maximum and minimum degrees, $u \sim v$ for adjacency, dist for the distance, $\mathrm{ecc}(v) = \max_w \mathrm{dist}(v,w)$, $\mathrm{rad} = \min_v \mathrm{ecc}(v)$ for the radius and $\mathrm{diam} = \max_v \mathrm{ecc}(v)$ for the diameter.

- A is the adjacency matrix, with eigenvalues $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$ (counted with multiplicity) and rank $\mathrm{rk}(A)$.
- $\mathcal{L} = D - A$ is the Laplacian, with eigenvalues $\mu_1 \geq \mu_2 \geq \cdots \geq \mu_n = 0$.
- $R = \sum_{uv \in E} (d_u d_v)^{-1/2}$ is the Randić index.

We shall freely use the following classical facts.

(F1) *Rayleigh*: $\lambda_1 = \max_{x \neq 0} x^\top A x / x^\top x$. *Courant–Fischer*: for a symmetric matrix M with eigenvalues $\nu_1 \geq \nu_2 \geq \cdots$, we have $\nu_2 = \max_{\dim U = 2} \min_{z \in U \setminus \{0\}} z^\top M z / z^\top z$.

(F2) *Perron–Frobenius*: since G is connected, $\lambda_1 > 0$ is simple and has an eigenvector with all coordinates strictly positive.

(F3) *Traces*: $\sum_i \lambda_i = \mathrm{tr}\, A = 0$, $\sum_i \lambda_i^2 = \mathrm{tr}\, A^2 = 2m$ and $\sum_i \mu_i = 2m$.

(F4) *Quadratic form*: $z^\top \mathcal{L} z = \sum_{uv \in E} (z_u - z_v)^2$.

(F5) $\lambda_1 \leq \Delta \leq n - 1$ and $m \geq n - 1$, with equality if and only if G is a tree.

(F6) *Smith* [8]: if G is complete, then $\lambda_2 = -1$, $\lambda_1 = n - 1$, $\mathrm{rk}(A) = n$ and $R = n/2$; otherwise G contains an induced P_3 and, by interlacing, $\lambda_2 \geq \lambda_2(P_3) = 0$.

Status of the results. The label *(verified in `Lean`)* means that the statement is a `Lean` 4 theorem proved for every connected graph `SimpleGraph (Fin n)`. This theorem depends only on the standard axioms (`propext`, `Classical.choice`, `Quot.sound`) and contains no `sorry`. The proofs below are mathematical transcriptions of the formal proofs.

2 Discovered and proved lemmas

The lemmas of this section were *discovered* automatically. On a database of 281,672 graphs (all connected graphs with at most 9 vertices, plus extremal families), we searched for linear inequalities between two invariants that ruled out the abstract countermodels returned by z3. They were then proved.

2.1 Laplacian and adjacency: $\mu_2 \geq \lambda_2 + 1$

Theorem 1 (*(verified in Lean)*). *For every connected graph, $\mu_2 \geq \lambda_2 + 1$. Equality holds for instance for stars ($\lambda_2 = 0$, $\mu_2 = 1$).*

Proof. If $n = 2$, then $G = K_2$, $\lambda_2 = -1$ and $\mu_2 = 0$. Assume $n \geq 3$.

Case $\lambda_2 \leq 0$. It suffices to show $\mu_2 \geq 1$. Since G is connected with $n \geq 3$, it has a vertex c with two neighbours a, b. Let H be the spanning graph with edges ca, cb. The spectrum of $\mathcal{L}_H$ is $\{3, 1, 0, \ldots, 0\}$. Since $\mathcal{L}_G = \mathcal{L}_H + \mathcal{L}_{G \setminus E(H)}$ and the second term is positive semidefinite, Courant–Fischer gives $\mu_2(G) \geq \mu_2(H) = 1$.

Case $\lambda_2 > 0$. Let $x > 0$ be the Perron vector (F2) and y an eigenvector of λ_2 orthogonal to x. Since $\sum_v x_v y_v = 0$ with $x > 0$ and $y \neq 0$, the vector y takes both strictly positive and strictly negative values.

Let u be a vertex where y is maximal ($y_u > 0$), $P = \{v \sim u : y_v > 0\}$ and $d^+ = |P|$. Then

$$\lambda_2 \, y_u = \sum_{v \sim u} y_v \leq \sum_{v \in P} y_v \leq d^+ \, y_u, \qquad \text{so } d^+ \geq \lambda_2.$$

Likewise, if w attains the minimum of y ($y_w < 0$), $N = \{v \sim w : y_v < 0\}$ and $d^- = |N|$, we get $\lambda_2 y_w \geq d^- y_w$, hence $d^- \geq \lambda_2$. In particular $d^{\pm} \geq 1$.

The sets $\{u\} \cup P$ (where $y > 0$) and $\{w\} \cup N$ (where $y < 0$) are disjoint. Set

$$f_1 = d^+ e_u - \mathbf{1}_P, \qquad f_2 = d^- e_w - \mathbf{1}_N.$$

These vectors have disjoint supports, so they span a plane U. For $z = a f_1 + b f_2$, we keep in (F4) only the edges uj ($j \in P$) and wj ($j \in N$). These two families of edges are disjoint, because $w \notin P$ and $u \notin N$. On the edge uj we have $z_u - z_j = a(d^+ + 1)$, hence

$$z^\top \mathcal{L} z \;\geq\; d^+(d^+ + 1)^2 a^2 + d^-(d^- + 1)^2 b^2, \qquad z^\top z = d^+(d^+ + 1) a^2 + d^-(d^- + 1) b^2.$$

The Rayleigh quotient is therefore at least $\min(d^+, d^-) + 1$ on U. Courant–Fischer then gives $\mu_2 \geq \min(d^+, d^-) + 1 \geq \lambda_2 + 1$. □

Remark. The proof actually establishes the stronger inequality $\mu_2 \geq \min(d^+, d^-) + 1$, where $d^{\pm}$ are the "signed degrees" at the extremities of an eigenvector of λ_2. A quick literature search did not find the inequality $\mu_2 \geq \lambda_2 + 1$. A more thorough check is needed before claiming that it is new.

2.2 Randić index and spectral radius

The following two classical results serve as tools. They were formalised, which made the corollaries almost immediate.

Theorem 2 (Favaron–Mahéo–Saclé, *(verified in Lean)*). *$m \leq \lambda_1 R$.*

Proof. With $x_v = \sqrt{d_v}$, we have $x^\top x = 2m$ and $x^\top A x = 2 \sum_{uv \in E} \sqrt{d_u d_v}$. Rayleigh gives $S := \sum_{uv \in E} \sqrt{d_u d_v} \leq m \lambda_1$. By Cauchy–Schwarz,

$$m^2 = \Big(\sum_{uv \in E} (d_u d_v)^{1/4} (d_u d_v)^{-1/4} \Big)^2 \leq S \cdot R \leq m \lambda_1 R.$$ □

Theorem 3 (Hong, *(verified in Lean)*). $\lambda_1^2 \le 2m - n + 1$.

Proof. Let x be a unit eigenvector of λ_1 and $P_v = \sum_{u\sim v} x_u^2$. By Cauchy–Schwarz, $\lambda_1^2 x_v^2 = (\sum_{u\sim v} x_u)^2 \le d_v P_v$. Since v is not its own neighbour, $P_v \le 1 - x_v^2$. Since $d_v \ge 1$, we get $d_v P_v \le P_v + (d_v - 1)(1 - x_v^2)$. Summing over v, with $\sum_v P_v = \sum_v d_v x_v^2$:

$$\lambda_1^2 \le \sum_v d_v x_v^2 + \sum_v (d_v - 1) - \sum_v d_v x_v^2 + \sum_v x_v^2 = 2m - n + 1. \qquad \square$$

Corollary 4 (*(verified in Lean)*). $R \ge \dfrac{\lambda_1 + 1}{2}$.

Proof. By Hong and $\lambda_1 \le n - 1$: $\lambda_1(\lambda_1 + 1) \le (2m - n + 1) + (n - 1) = 2m$. By Theorem 2, $\lambda_1 R \ge m \ge \lambda_1(\lambda_1 + 1)/2$. Divide by $\lambda_1 > 0$. $\square$

Corollary 5 (Bollobás–Erdős and upper bound, *(verified in Lean)*). $\sqrt{n-1} \le R \le n/2$.

Proof. Upper bound. By the AM–GM inequality, $(d_u d_v)^{-1/2} \le \frac{1}{2}(1/d_u + 1/d_v)$. Summing over the edges, each vertex v appears d_v times, so $R \le \frac{1}{2}\sum_v d_v \cdot \frac{1}{d_v} = n/2$.

Lower bound. Set $q = 2m - n + 1 > 0$. Theorems 2 and 3 give $m^2 \le \lambda_1^2 R^2 \le qR^2$. Moreover, $m^2 - (n-1)q = (m - n + 1)^2 \ge 0$. Hence $(n-1)q \le m^2 \le qR^2$, and we divide by q. $\square$

Theorem 6 (*(verified in Lean)*). $R \ge \dfrac{2}{3}\lambda_2 + \sqrt{2}$.

Proof. Write $L = \lambda_1$, $t = \lambda_2$ and $s = \sqrt{2}$. By Theorem 2, it suffices to prove

$$L\Big(\tfrac{2}{3}t + s\Big) \le m, \tag{1}$$

because then $LR \ge m \ge L(\frac{2}{3}t + s)$, and we divide by $L > 0$.

Case 1: G is not a tree. Then $n \ge 3$ and $m \ge n$. Set $k = n - 2 \ge 1$. By (F3), $\sum_{i\ge 3} \lambda_i = -(L + t)$ and $\sum_{i\ge 3} \lambda_i^2 = 2m - L^2 - t^2$. Cauchy–Schwarz on the k eigenvalues $\lambda_3, \dots, \lambda_n$ gives

$$(L + t)^2 \le k\,(2m - L^2 - t^2). \tag{2}$$

One then checks the polynomial identity

$$k\Big(m - L(\tfrac{2}{3}t + s)\Big) = \tfrac{7}{20}\Big(2mk - k(L^2 + t^2) - (L + t)^2\Big) + \tfrac{3}{10}k(m - (k + 2)) + Q,$$

where $Q = a\,t^2 + b\,Lt + a\,L^2 - s\,kL + \frac{3}{10}k(k+2)$, with $a = \frac{7}{20}(k+1)$ and $b = \frac{7}{10} - \frac{2k}{3}$. The first two terms are nonnegative by (2) and because $m \ge n = k + 2$. It remains to see that $Q \ge 0$. With $P = \frac{41}{900}k + \frac{287}{150} > 0$, we have

$$4aQ = (2at + bL)^2 + k\,W, \qquad W = PL^2 - \tfrac{7}{5}(k+1)sL + \tfrac{21}{50}(k+1)(k+2),$$
$$4PW = \big(2PL - \tfrac{7}{5}(k+1)s\big)^2 + \tfrac{14}{25}(k+1)\,V, \qquad V = \tfrac{41k^2 - 296k + 1344}{300}.$$

The discriminant of $41k^2 - 296k + 1344$ is $296^2 - 4 \cdot 41 \cdot 1344 < 0$, so $V > 0$. Hence $W \ge 0$, then $Q \ge 0$, which proves (1).

Case 2: G is a tree, so $m = n - 1$. A tree is bipartite. Let $\sigma_v = \pm 1$ according to the class of v. If $Ax = \lambda x$, then the vector $(\sigma_v x_v)_v$ is an eigenvector for $-\lambda$: the spectrum is symmetric.

- If $t > 0$, then $L > t > 0$, and $L, t, -t, -L$ are four distinct eigenvalues. By (F3), $2L^2 + 2t^2 \le 2m$, so $n \ge 4$ and $m \ge 3$. We conclude with the identity

  $$m - \tfrac{2}{3}Lt - sL = \tfrac{16}{25}(m - L^2 - t^2) + \tfrac{9}{25}(m - 3) + \tfrac{16}{25}\big(t - \tfrac{25}{48}L\big)^2 + \tfrac{1679}{3600}\big(L - \tfrac{1800\sqrt{2}}{1679}\big)^2 + \tfrac{333}{41975},$$

  all of whose terms are nonnegative.

- If $t \le 0$ and $n = 2$, then $G = K_2$ and $L(\frac{2}{3}t + s) = s - \frac{2}{3} < 1 = m$.
- If $t \le 0$ and $n \ge 3$, then L and $-L$ are two eigenvalues, so $2L^2 \le 2m$. Thus $L(\frac{2}{3}t + s) \le sL \le \sqrt{2m} \le m$, since $m \ge 2$. □

Remark. The certificate identities were found by a sum-of-squares search and checked symbolically; they are also checked by Lean.

2.3 Distances and rank of the adjacency matrix

Lemma 7. *If $v_1, \dots, v_p$ is an induced path of G, then* $\mathrm{rk}(A) \ge p - 1$. *If moreover $p \ge 4$, then* $\mathrm{rk}(A) \ge 4$.

Proof. Consider the submatrix of A with rows $v_1, \dots, v_{p-1}$ and columns $v_2, \dots, v_p$. Its (i, j) entry is $A_{v_i, v_{j+1}}$.

- It equals 1 if $j = i$.
- It is zero if $j > i$, because v_i and v_{j+1} are not adjacent ($j + 1 \ge i + 2$ and the path is induced).

The submatrix is therefore lower triangular with unit diagonal, of determinant 1, and $\mathrm{rk}(A) \ge p - 1$. If $p \ge 4$, the vertices $v_1, \dots, v_4$ induce P_4, whose adjacency matrix has determinant 1. An invertible principal submatrix of order 4 gives $\mathrm{rk}(A) \ge 4$. □

Corollary 8 (*(verified in Lean)*)**.** $\mathrm{rk}(A) \ge \mathrm{diam}$, *and* $\mathrm{rk}(A) \ge 4$ *as soon as* $\mathrm{diam} \ge 3$.

Proof. A geodesic between two vertices at distance diam is an induced path with diam $+1$ vertices. Apply Lemma 7. □

Theorem 9 (*(verified in Lean)*)**.** $\mathrm{rad} \le \dfrac{\mathrm{rk}(A)}{2} + 1$.

Proof. If $\mathrm{rad} \le 1$, this is immediate since $\mathrm{rk}(A) \ge 2$. Otherwise, we use the induced path theorem of Erdős, Saks and Sós (in F. Chung's proof, which is also formalised): a connected graph of radius r contains an induced path with at least $2r - 1$ vertices. Lemma 7 then gives $\mathrm{rk}(A) \ge 2r - 2$. □

Remark. This theorem relates a metric quantity, the radius, to a purely algebraic quantity, the rank. It is attained (zero margin) on the graph database.

3 The ten most useful lemmas

A proof found by the prover is a *minimal* (for inclusion) set of library lemmas from which the conjecture follows by nonlinear real arithmetic. We measure the usefulness of a lemma by the number of these proofs that use it, among the 117,390 conjectures proved without any unproved hypothesis. The hypothesis $n \ge 2$ of the setting, which appears in 7,670 proofs, is not counted as a lemma.

	Name	Statement	Proofs	Lean
1	`rad_dia_pos`	$\mathrm{rad} \geq 1$, $\mathrm{diam} \geq 1$	40,630	✓
2	`rank_range`	$2 \leq \mathrm{rk}(A) \leq n$	40,451	✓
3	`rad_dia`	$\mathrm{rad} \leq \mathrm{diam} \leq 2\,\mathrm{rad}$	40,124	✓
4	`rank_dia`	$\mathrm{rk}(A) \geq \mathrm{diam}$, and $\mathrm{rk}(A) \geq 4$ if $\mathrm{diam} \geq 3$	28,147	✓
5	`randic_n`	$\sqrt{n-1} \leq R \leq n/2$	24,731	✓
6	`R_pos`	$R > 0$	20,614	✓
7	`yang_lu`	$R \geq \mathrm{diam}/2 + \sqrt{2} - 1$	12,949	✓
8	`rad_n`	$2\,\mathrm{rad} \leq n$	9,545	
9	`dia_n`	$\mathrm{diam} \leq n - 1$	8,374	✓
10	`m_bounds`	$n - 1 \leq m \leq n(n-1)/2$	6,524	✓

The most used lemmas are elementary. They mostly serve to rule out degenerate cases (small graphs, zero values, signs) in inequalities whose core lies elsewhere. Among the deep lemmas, the theorem of Yang and Lu clearly dominates: it occurs in 12,949 proofs, and in the proof of 2,301 of the 2,906 theorems that require a research-level result (Section 5). Next come $R \geq (\lambda_1 + 1)/2$ (Corollary 4, 550 proofs), $R \geq \frac{2}{3}\lambda_2 + \sqrt{2}$ (Theorem 6, 544), $\mathrm{rad} \leq \mathrm{rk}(A)/2 + 1$ (Theorem 9, 478) and $\mu_2 \geq \lambda_2 + 1$ (Theorem 1, 105).

Lemma 10 (`rad_dia_pos`, *(verified in Lean)*)**.** $\mathrm{rad} \geq 1$ *and* $\mathrm{diam} \geq 1$.

Proof. Since $n \geq 2$, every vertex v has another vertex w, and $\mathrm{dist}(v, w) \geq 1$. Hence $\mathrm{ecc}(v) \geq 1$ for every v. □

Lemma 11 (`rank_range`, *(verified in Lean)*)**.** $2 \leq \mathrm{rk}(A) \leq n$.

Proof. The upper bound is obvious. Since G is connected with $n \geq 2$, it has an edge uv. The principal submatrix of A on $\{u, v\}$ is $\left(\begin{smallmatrix} 0 & 1 \\ 1 & 0 \end{smallmatrix}\right)$, of determinant -1. Hence $\mathrm{rk}(A) \geq 2$. □

Lemma 12 (`rad_dia`, *(verified in Lean)*)**.** $\mathrm{rad} \leq \mathrm{diam} \leq 2\,\mathrm{rad}$.

Proof. rad is a minimum and diam a maximum of the same eccentricities. If c attains the radius, the triangle inequality gives $\mathrm{dist}(u, v) \leq \mathrm{dist}(u, c) + \mathrm{dist}(c, v) \leq 2\,\mathrm{rad}$ for all u, v. □

Lemma 13 (`rank_dia`, *(verified in Lean)*)**.** $\mathrm{rk}(A) \geq \mathrm{diam}$*, and* $\mathrm{rk}(A) \geq 4$ *if* $\mathrm{diam} \geq 3$.

Proof. This is Corollary 8. □

Lemma 14 (`randic_n`, *(verified in Lean)*)**.** $\sqrt{n-1} \leq R \leq n/2$.

Proof. This is Corollary 5. The lower bound is due to Bollobás and Erdős [4]; the shorter proof given here is discussed in Section 6. □

Lemma 15 (`R_pos`, *(verified in Lean)*)**.** $R > 0$.

Proof. G has at least one edge and every edge contributes $(d_u d_v)^{-1/2} > 0$. □

Lemma 16 (`yang_lu`, Yang–Lu [6], *(verified in Lean)*)**.** $R \geq \dfrac{\mathrm{diam}}{2} + \sqrt{2} - 1$.

Proof. See Section 6, which describes the formalised proof and how it differs from the one in the paper. □

Lemma 17 (`rad_n`)**.** $2\,\mathrm{rad} \leq n$.

Proof. Let T be a spanning tree of G; distances in T bound those of G from above. Let $P = p_0 p_1 \cdots p_D$ be a longest path of T and $c = p_k$ with $k = \lceil D/2 \rceil$. Let w be a vertex, p_j the first vertex of P met by the path of T from w to c, and $h = \mathrm{dist}_T(w, p_j)$.

- If $j \le k$, the path $w \to p_j \to p_D$ has length $h + D - j \le D$, so $h \le j$ and $\mathrm{dist}_T(w, c) = h + k - j \le k$.
- If $j \ge k$, the path $w \to p_j \to p_0$ has length $h+j \le D$, so $\mathrm{dist}_T(w, c) = h+j-k \le D-k \le k$.

Thus $\mathrm{rad} \le \mathrm{ecc}_G(c) \le \mathrm{ecc}_T(c) \le \lceil D/2 \rceil \le \lceil (n-1)/2 \rceil \le n/2$. □

Lemma 18 (`dia_n`, *(verified in Lean)*)**.** $\mathrm{diam} \le n - 1$.

Proof. A geodesic of length diam goes through diam $+1$ distinct vertices. □

Lemma 19 (`m_bounds`, *(verified in Lean)*)**.** $n - 1 \le m \le n(n-1)/2$.

Proof. A connected graph contains a spanning tree, which has $n - 1$ edges. A simple graph has at most $\binom{n}{2}$ edges. □

4 Proved conjectures of the corpus

Here is a selection of conjectures of the corpus, identified by their number, that follow from the previous results.

Proposition 20. *For every connected graph:*

(a) ***#20832****:* $\lambda_2 \le \mu_2 - 1$*, attained* (verified in Lean)*;*

(b) ***#20821****:* $\lambda_2 \le R^2 - 2$*, attained for* K_2*;*

(c) ***#18099****:* $n - 1 \le R\lambda_1$*, and* ***#764799****:* $\mathrm{diam} \le R\lambda_1$ (verified in Lean)*;*

(d) ***#127304****:* $\mu_2 \le R\lambda_1$*;*

(e) ***#759987****:* $\mathrm{diam} \le 2R$*, and* ***#760654****:* $\mathrm{diam} \le \mathrm{rad} + R$ (verified in Lean)*;*

(f) ***#1362170****:* $\lambda_1 + \mathrm{rad} \le n$*, attained for* K_n *and* C_4 (verified in Lean).

Proof. (a) This is Theorem 1.

(b) If $\lambda_2 \ge 0$, Theorem 6 gives $R^2 \ge (\frac{2}{3}\lambda_2 + \sqrt{2})^2 = \frac{4}{9}\lambda_2^2 + \frac{4\sqrt{2}}{3}\lambda_2 + 2 \ge \lambda_2 + 2$. If $\lambda_2 < 0$, the graph is complete (Smith), so $\lambda_2 = -1$ and $R = n/2 \ge 1$.

(c) Use Theorem 2 and $\mathrm{diam} \le n - 1 \le m$.

(d) We have $2\mu_2 \le \mu_1 + \mu_2 \le \sum_i \mu_i = 2m$, so $\mu_2 \le m \le \lambda_1 R$.

(e) Yang and Lu proved $R \ge \frac{1}{2}\,\mathrm{diam} + \sqrt{2} - 1$, which gives $\mathrm{diam} \le 2R$. With $\mathrm{diam} \le 2\,\mathrm{rad}$, we get $\mathrm{diam} \le \frac{1}{2}\,\mathrm{diam} + \mathrm{rad} \le R + \mathrm{rad}$. The theorem of Yang and Lu is itself formalised in Lean (`Spectral.L_yang_lu`, about 3,500 lines, without `sorry`). The formal proof keeps the minimal counterexample strategy of the paper, but replaces its second half with a different argument. A strong induction on the number of edges first shows that a non-essential vertex of a minimal counterexample has degree at least 9. It then shows that degrees grow exponentially with the distance to the set of essential vertices. Finally, a diametral geodesic is cut into pieces, each of weight at least $\ell/2 + 1$ thanks to a local version of Lemma 1 of the paper. The formalisation uncovered several gaps in the paper: a false equivalence about essential vertices (counterexample: C_5 plus a pendant vertex), an unproved diameter inequality in Corollary 4, hypotheses of Lemma 6 that are never checked, an incomplete block structure, and index errors in Claims A and B.

(f) First $\lambda_1 \le \Delta$: if $Ax = \lambda_1 x$ and u maximises $|x_u|$, then $\lambda_1 |x_u| = |\sum_{v \sim u} x_v| \le d_u |x_u| \le \Delta |x_u|$. Next, let v be a vertex of degree Δ, w any vertex, $k = \mathrm{dist}(v, w)$ and $v = p_0, p_1, \ldots, p_k = w$ a geodesic. The vertices $p_2, \ldots, p_k$ are pairwise distinct, distinct from v and not adjacent to v (since $\mathrm{dist}(v, p_i) = i \ge 2$). The sets $\{v\}$, $N(v)$ and $\{p_2, \ldots, p_k\}$ are therefore disjoint, so $1 + \Delta + (k-1) \le n$. Thus $\mathrm{rad} \le \mathrm{ecc}(v) \le n - \Delta$, and $\lambda_1 + \mathrm{rad} \le \Delta + (n - \Delta) = n$. □

Remark. (f) remained "open" for the solver only because its library did not contain the lemma $\mathrm{rad} \le n - \Delta$. (c) and (d) were proved only after the theorem of Favaron–Mahéo–Saclé was added. This illustrates how enriching the library brings down entire families of conjectures.

5 The ten hardest theorems

Criterion. The internal measures of the prover (number of lemmas, computing time) mostly reflect case distinctions on small graphs. The statements that maximise these measures are nested powers such as $\mathrm{rad} \le R^{2^{\sqrt{\sqrt{\mathrm{rk}(A)}}}}$, which often have a three-line human proof. We therefore used two more meaningful criteria.

(i) *Depth*: the theorem does not follow from the elementary and classical lemmas of the library alone (levels 0 to 2). We checked this by rerunning z3 without the research-level lemmas: it then finds a genuine countermodel. Among the 6,651 research-level theorems verified in Lean, 3,679 can in fact be proved without any deep result. There remain 2,906 that require one of Theorems 1, 6, 9 or the theorem of Yang and Lu.

(ii) *Optimality*: the inequality is attained (with equality) on at least one graph.

Among the theorems that satisfy (i) and (ii), we chose the following ten, avoiding variants of the same statement. The first three are verified in Lean in abstract form: the implication from the lemmas is formal, but Smith's fact (F6) that they use is not yet formalised. The other seven are verified end to end.

	Statement	Deep result	Equality
#118503	$\lambda_2 \le 2(R - \sqrt{2})$	Th. 6	P_3
#709846	$\lambda_2 \le \mathrm{rk}(A)\,(R - \sqrt{2})$	Th. 6	P_3
#2630815	$\lambda_2 \le \mathrm{diam}\,(R - \sqrt{2})$	Th. 6	P_3
#721942	$\lambda_2 \le 2\mu_2 - 1$	Th. 1	K_2
#1935857	$\mathrm{diam} \le R^{3/\sqrt{2}}$	Yang–Lu	K_2
#1935639	$\mathrm{diam} \le R\,(\sqrt{2} + \log \mathrm{rad})$	Yang–Lu	P_3
#869836	$\mathrm{rad} \le \mathrm{rk}(A) - \log(\mathrm{rad}\,/2)$	Th. 9	C_4
#1396267	$\mathrm{rad} \le 2^{\mathrm{rk}(A) - \mathrm{diam}\,/2}$	Th. 9	C_4
#1356167	$\mathrm{rad} \le 1 + \mathrm{rk}(A) - \mathrm{rad}\,/2$	Th. 9	C_4
#1454913	$\mathrm{rad} \le \mathrm{rk}(A) - \sqrt{\mathrm{rad}} + \sqrt{2}$	Th. 9	C_4

In the proofs below, we write $t = \lambda_2$, $k = \mathrm{rk}(A)$, $D = \mathrm{diam}$ and $r = \mathrm{rad}$.

Theorem 21 (#118503)**.** *$\lambda_2 \le 2(R - \sqrt{2})$, with equality for P_3.*

Proof. If $t \ge 0$, Theorem 6 gives $R \ge \frac{2}{3}t + \sqrt{2} \ge \frac{1}{2}t + \sqrt{2}$. Otherwise G is complete (F6), $t = -1$ and $R = n/2 \ge 1 > \sqrt{2} - \frac{1}{2}$. For P_3: $t = 0$ and $R = \sqrt{2}$. □

Theorem 22 (#709846)**.** *$\lambda_2 \le \mathrm{rk}(A)\,(R - \sqrt{2})$, with equality for P_3.*

Proof. If $G = K_n$ (F6): $k(R - \sqrt{2}) = n(\frac{n}{2} - \sqrt{2})$ equals $2 - 2\sqrt{2} > -1$ for $n = 2$ and is positive for $n \ge 3$; it is therefore at least $t = -1$. Otherwise $t \ge 0$, and Theorem 6 gives $R - \sqrt{2} \ge \frac{2}{3}t \ge 0$. With $k \ge 2$: $k(R - \sqrt{2}) \ge \frac{4}{3}t \ge t$. For P_3: $k = 2$, $R = \sqrt{2}$, $t = 0$. □

Theorem 23 (#2630815)**.** *$\lambda_2 \le \mathrm{diam}\,(R - \sqrt{2})$, with equality for P_3.*

Proof. If G is complete, $D = 1$, $t = -1$ and $R - \sqrt{2} \ge 1 - \sqrt{2} > -1$. Otherwise $D \ge 2$, $t \ge 0$ and $R - \sqrt{2} \ge \frac{2}{3}t \ge 0$, hence $D(R - \sqrt{2}) \ge \frac{4}{3}t \ge t$. □

Theorem 24 (#721942, *(verified in Lean)*)**.** *$\lambda_2 \le 2\mu_2 - 1$, with equality for K_2.*

Proof. By Theorem 1 and $\mu_2 \geq 0$: $2\mu_2 - 1 = (\mu_2 - 1) + \mu_2 \geq t$. For K_2: $\mu_2 = 0$ and $t = -1$. □

Theorem 25 (#1935857, *(verified in Lean)*)**.** $\operatorname{diam} \leq R^{3/\sqrt{2}}$*, with equality for* K_2*.*

Proof. By Corollary 5, $R \geq \sqrt{n-1} \geq 1$; since $3/\sqrt{2} > 2$, we have $R^{3/\sqrt{2}} \geq R^2 \geq 1$. This settles the case $D = 1$. If $D \geq 2$, Yang–Lu gives $R^2 \geq (\frac{D}{2} + \sqrt{2} - 1)^2$. Now $\varphi(D) = (\frac{D}{2} + \sqrt{2} - 1)^2 - D$ vanishes at $D = 2$ and $\varphi'(D) = \frac{D}{2} + \sqrt{2} - 2 > 0$ for $D \geq 2$, so $\varphi(D) \geq 0$. □

Theorem 26 (#1935639, *(verified in Lean)*)**.** $\operatorname{diam} \leq R\,(\sqrt{2} + \log \operatorname{rad})$*, with equality for* P_3*.*

Proof. We have $\log r \geq 0$. If $D \leq 2$, Yang–Lu gives $\sqrt{2}R \geq \frac{D}{\sqrt{2}} + 2 - \sqrt{2}$, and this quantity is $\geq D$ if and only if $D \leq 2$. If $D \geq 3$, then $r \geq D/2 > 1$, so $r \geq 2$ (lemma `rad_dia`). Consequently $\sqrt{2} + \log r \geq \sqrt{2} + \log 2 > 2.1$, and $R > D/2$ by Yang–Lu, hence $R(\sqrt{2} + \log r) > 1.05\,D$. For P_3: $D = 2$, $r = 1$, $R = \sqrt{2}$. □

Theorem 27 (#869836, *(verified in Lean)*)**.** $\operatorname{rad} \leq \operatorname{rk}(A) - \log(\operatorname{rad}/2)$*, with equality for* C_4*.*

Proof. If $r \leq 2$, then $\log(r/2) \leq 0$ and $k \geq 2 \geq r$. If $r \geq 2$, Theorem 9 gives $k \geq 2r - 2$, and $\log(r/2) \leq \frac{r}{2} - 1 \leq r - 2$. Hence $k - \log(r/2) \geq r$. For C_4: $r = k = 2$. □

Theorem 28 (#1396267, *(verified in Lean)*)**.** $\operatorname{rad} \leq 2^{\operatorname{rk}(A) - \operatorname{diam}/2}$*, with equality for* C_4*.*

Proof. Since $D \leq k$ (lemma `rank_dia`), we have $k - \frac{D}{2} \geq \frac{k}{2} \geq r - 1$ by Theorem 9. Hence $2^{k - D/2} \geq 2^{r-1} \geq r$ for every integer $r \geq 1$. For C_4: $k = D = r = 2$. □

Theorem 29 (#1356167, *(verified in Lean)*)**.** $\operatorname{rad} \leq 1 + \operatorname{rk}(A) - \operatorname{rad}/2$*, with equality for* C_4*.*

Proof. This reads $\frac{3}{2}r \leq k + 1$. If $r = 1$, use $k \geq 2$. If $r \geq 2$, Theorem 9 gives $k + 1 \geq 2r - 1 \geq \frac{3}{2}r$. □

Theorem 30 (#1454913, *(verified in Lean)*)**.** $\operatorname{rad} \leq \operatorname{rk}(A) - \sqrt{\operatorname{rad}} + \sqrt{2}$*, with equality for* C_4*.*

Proof. If $r = 1$, $k \geq 2 > 2 - \sqrt{2}$. If $r \geq 2$, $k \geq 2r - 2$, and it suffices that $r - \sqrt{r} \geq 2 - \sqrt{2}$. The function $x \mapsto x - \sqrt{x}$ is increasing on $[1, \infty)$ and equals $2 - \sqrt{2}$ at $x = 2$. □

Remark. Once the deep results are available, these proofs are short. The difficulty lies in those results themselves: the formal proof of Yang–Lu has about 3,500 lines, and $\mu_2 \geq \lambda_2 + 1$ needs a two-dimensional Courant–Fischer argument. This is also why enriching the library brings down entire families of conjectures.

6 Simpler proofs of known results

The formalisation led us to replace some proofs from the literature with arguments that are shorter, or easier to make rigorous. We do not know whether these arguments already appear in the literature.

6.1 The lower bound of Bollobás and Erdős

Bollobás and Erdős [4] proved that $R \geq \sqrt{n-1}$ for every connected graph, with equality for stars. Their proof determines the minimum by an extremal analysis. The proof of Corollary 5 uses only two classical spectral inequalities and fits in three lines.

Short proof. Set $q = 2m - n + 1 > 0$. Favaron–Mahéo–Saclé (Theorem 2) and Hong (Theorem 3) give $m^2 \leq \lambda_1^2 R^2 \leq qR^2$. On the other hand, $m^2 - (n-1)q = (m - n + 1)^2 \geq 0$. Hence $(n-1)q \leq qR^2$. □

The equality case can be read off the proof: it requires $m = n - 1$ (a tree) and $\lambda_1^2 = 2m - n + 1 = n - 1$, which characterises the star. This proof was not written by a human. The agent in charge of formalising the lemma reused the two theorems the library already contained, instead of following the original proof. Lemmas formalised earlier thus make later ones almost free.

6.2 The theorem of Yang and Lu

Yang and Lu [6] prove $R \geq \operatorname{diam}/2 + \sqrt{2} - 1$ by considering a minimal counterexample. The first half of their proof (vertex deletions, "essential" vertices) is kept. The second half, which relies on Lemmas 4, 6 and 7, a block structure read off figures, and Claims A and B, is replaced with a shorter argument that needs no analysis of figures. Here are its steps; we write $c = \sqrt{2} - 1$ and $f(G) = R - \operatorname{diam}/2$.

1. *Weak deletion* (replaces Lemma 2 and its orientations). Let v be a vertex, and d' the degrees in $G - v$. If, for every neighbour u of v of degree $d_u \geq 2$,

$$\textstyle\sum_{x \in N(u) \setminus v} 1/\sqrt{d'_x} \leq 2(d_u - 1)/\sqrt{d_v},$$

   then $R(G - v) \leq R(G)$. The proof is a simple term-by-term comparison, using the numerical inequality $2(d-1)(1/\sqrt{d-1} - 1/\sqrt{d}) \leq 1/\sqrt{d}$. This condition is automatically satisfied if $d_v \leq 4$.
2. *Minimal counterexample.* We choose a counterexample G with as few edges as possible, on a fixed vertex set (a deleted vertex stays isolated, which changes neither R nor diam). All reductions are then spanning subgraphs, and a single induction on the number of edges suffices. We show that the essential vertices lie on a diametral geodesic P, that the edges between essential vertices are bridges, that $G - e$ has only two components for an inner essential edge e, and that the inner essential vertices have degree at least 3.
3. *Degree of the non-essential vertices* of G (replaces the five cases of Lemma 3). Every non-essential vertex has degree at least 9. Indeed, a vertex of minimum degree $\delta \leq 8$ would satisfy the weak deletion condition, thanks to three inequalities that are affine in d_u and checked at their smallest value.
4. *Exponential growth* (replaces Claim A and its layers). By strong induction on the distance r to the set of essential vertices, a vertex at distance $\geq r$ has degree $\geq g(r)$, where $g(1) = 9$, $g(2) = 19$ and $g(r) = 4g(r-2) - 3$, so that $g(r) \geq 2^{r+2} + 1$.
5. *Cutting into pieces* (replaces the block structure and Claim B). The inner essential vertices cut P into segments of lengths ℓ_i, with $\sum \ell_i = \operatorname{diam}$. To each segment we associate the set I_i of non-essential vertices joined to it without going through an essential vertex. These sets are disjoint, and two consecutive segments are never both empty. A local version of Lemma 1 of the paper,

$$\textstyle\sum_{xy \in E,\ x,y \in I} \frac{1}{\sqrt{d_x d_y}} \ \geq\ \frac{1}{2\sqrt{\Delta_I}} \sum_{x \in I} \frac{|N(x) \cap I|}{\sqrt{d_x}},$$

   and step 4 give each nonempty piece a weight of at least $\ell_i/2 + 1 + c$. An elementary count concludes: $R \geq \sum_i \ell_i/2 + c = \operatorname{diam}/2 + c$, which contradicts $f(G) < c$.

The formalisation also uncovered several gaps in the paper.

- The equivalence claimed for essential vertices is false (counterexample: C_5 plus a pendant vertex). Only the true implication is used.
- Corollary 4 assumes $\operatorname{diam}(G - uv) \leq \operatorname{diam}(G) + 1$ without justification. We use instead $\operatorname{diam}(G - uv) \geq \operatorname{diam}(G) + 1$, which directly contradicts minimality.
- The hypotheses of Lemma 6 are never checked.
- The "pendant" components of $G - e$ are ignored by the block structure. They are handled by an additional deletion lemma.

- The indices in Claims A and B are wrong, and the final inequality, "easy to check by computation", is not proved.

The complete plan, with all the numerical inequalities, is included in the archive, in the file `documents/yang_lu_plan_de_preuve.md`.

7 The most interesting unrefuted conjectures

All the conjectures below were tested:

- on all connected graphs with at most 9 vertices;
- on structured families up to $n = 2000$ (complete k-ary trees, hypercubes, random regular graphs, grids, spiders, dumbbells…);
- on closed-form families (paths, cycles, stars, K_n, $K_{a,b}$) up to $n = 10^{15}$.

Conjecture 1 and Conjecture 7 were also tested on the 11.7 million connected graphs with 10 vertices.

These tests are not superfluous. Among the 16,630 conjectures that the prover left open, 2,480 turned out to be false on large graphs. For example, $R \le \mathrm{diam} \cdot \mu_1$ (#799714) is false for the complete binary tree with 255 vertices ($R \approx 116.1$ whereas $\mathrm{diam} \cdot \mu_1 \approx 78.6$).

Conjecture 1 (#643)**.** $R \ge \lambda_2 + 1$.

For two cliques K_k joined by an edge, $R - \lambda_2 - 1 \to 0$ (0.012 for $k = 80$). The conjecture would therefore be tight. Theorems 2 and 6 are not enough: the relaxation "traces + Hong + $m \ge n$" has numerical countermodels ($m = n = k + 2$, $\lambda_1 = \lambda_2 = \sqrt{k}$ for $k > 4$). On the other hand, the conjecture holds for all bipartite graphs, in particular for trees (Theorem 31): it remains open for non-bipartite graphs. We propose two strengthenings, with no counterexample up to $n = 10$:

$$2m \ge \lambda_1^2 + (\lambda_2 + 1)^2 \implies m \ge \lambda_1(\lambda_2 + 1) \implies R \ge \lambda_2 + 1.$$

The second implication follows from Theorem 2.

Theorem 31 (*(verified in Lean)*)**.** *If G is connected and bipartite (in particular if G is a tree), then $R \ge \lambda_2 + 1$.*

Proof. Write $L = \lambda_1$ and $t = \lambda_2$.

If $t \le 0$, Corollary 5 gives $R \ge \sqrt{n-1} \ge 1 \ge t + 1$.

If $t > 0$, we argue as in Case 2 of Theorem 6. The spectrum of a bipartite graph is symmetric. If x and y are orthogonal unit eigenvectors associated with L and t, and σ is the ± 1 signature of the two classes, then σx and σy are eigenvectors for $-L$ and $-t$. The four vectors $x, y, \sigma x, \sigma y$ are orthonormal: eigenvectors of eigenvalues of opposite signs are orthogonal, and multiplication by σ preserves the inner product. Hence $n \ge 4$, so $m \ge 3$. Bessel's inequality with (F3) gives $2L^2 + 2t^2 \le 2m$. We conclude with the identity

$$m - Lt - L = \tfrac{7}{10}(m - L^2 - t^2) + \tfrac{3}{10}(m - 3) + \tfrac{7}{10}\left(t - \tfrac{5}{7}L\right)^2 + \tfrac{12}{35}\left(L - \tfrac{35}{24}\right)^2 + \tfrac{41}{240},$$

all of whose terms are nonnegative. Hence $m \ge L(t+1)$, then $LR \ge m \ge L(t+1)$ by Theorem 2, and we divide by $L > 0$. □

Conjecture 2 (#20829)**.** $\lambda_2 \le \mu_1 - 2$.

It is asymptotically tight on paths ($\lambda_2 \to 2$, $\mu_1 \to 4$). Theorem 1 only gives $\lambda_2 \le \mu_2 - 1 \le \mu_1 - 1$.

Conjecture 3 (#501193)**.** $2R \le \mathrm{rk}(A) \cdot \lambda_1$.

Equality holds for *all* complete bipartite graphs $K_{a,b}$ ($\mathrm{rk}(A) = 2$, $\lambda_1 = R = \sqrt{ab}$). It is an elegant statement relating rank, spectral radius and Randić index. Note that $R \leq \lambda_1$ is false in general (paths), so the rank is indispensable.

Conjecture 4 (#112951)**.** $\lambda_1 \leq R + \mu_2$.

It mixes three spectra: adjacency, Laplacian and Randić. Equality holds for K_2.

Conjecture 5 (#239534)**.** $\mu_2 + \mathrm{rk}(A) \leq 2m$.

Equality holds for K_2, K_3 and P_4. For a tree, $\mathrm{rk}(A)$ is twice the size ν of a maximum matching, and the conjecture reads $2\nu + \mu_2 \leq 2n - 2$.

Conjecture 6 (unproved mined lemmas)**.** (i) $R^2 \geq \dfrac{m+1}{2}$. It improves Bollobás–Erdős ($R^2 \geq n - 1$) for dense graphs.
(ii) $\mu_2^2 \geq 4\lambda_2 + 1$, with equality for stars.
(iii) $\mu_1^2 \geq 4\lambda_2 + 8$, asymptotically tight on paths.
(iv) $R \geq \frac{2}{3}\sqrt{m} + \frac{1}{3}$ and $R \geq \frac{3}{2}\sqrt{\mathrm{rk}(A)} + 1 - \frac{3\sqrt{2}}{2}$, with equality for K_2.

Conjecture 7 (Caporossi–Hansen form)**.** $R \geq \mathrm{rad} + \sqrt{2} - \frac{3}{2}$.

Equality holds for paths of even order. Caporossi and Hansen proved it for trees. It follows from their conjecture $R \geq \mathrm{rad}$ (except for even paths) and implies Fajtlowicz's conjecture $R \geq \mathrm{rad} - 1$. It was rediscovered automatically by lemma mining. It is the most important from the point of view of the literature, but also the hardest. On its own, it would unlock more than 3,000 conditional proofs of the corpus.

8 Archive of the proofs

All the proofs are collected in the archive `preuves_spectrales.zip`, which contains:

- `index_conjectures.csv`: one line per conjecture of the corpus (305,755), with its status, its statement, the lemmas of its proof, the counterexample if any and, for the 79,505 verified proofs, the corresponding Lean file and theorem;
- `lemmes.csv`: the library of lemmas, with statement, reference, Lean theorem and number of proofs that use each lemma;
- `lean/`: the complete Lean 4 project (the `SpectralLean` library, including the proof of Yang–Lu) and the 3,650 proof files, with a script that recompiles them all;
- `documents/`: this report and the detailed plan of the proof of Yang–Lu;
- `prouveur/`: the code of the pipeline (refutation, proof by z3, lemma mining, Lean export).

The file names and column headers of the archive are in French (`LISEZMOI.md` is the README). For each verified conjecture #i, the theorem `g`i states the inequality for every connected simple graph on `Fin n`. To trust it, it suffices to read the definitions of the invariants (`SpectralLean/Invariants.lean`) and its statement.

Summary in figures

Conjectures in the corpus	305,755
Refuted	162,964
Proved without any unproved hypothesis	117,390
of which verified end to end in Lean	79,505
Proved under an empirical lemma or an open conjecture	3,798
Undefined on some graphs	7,453
Open and unrefuted (after tests on large graphs)	14,150

References

[1] E. Duhamel, T. Cazenave, Spectral Graph Theory Conjecture Generation, *AAAI 2026 Workshop AI for Research*; corpus: `https://www.lamsade.dauphine.fr/~cazenave/papers/Conjectures.zip`.

[2] O. Favaron, M. Mahéo, J.-F. Saclé, Some eigenvalue properties in graphs (conjectures of Graffiti II), *Discrete Math.* 111 (1993).

[3] Y. Hong, A bound on the spectral radius of graphs, *Linear Algebra Appl.* 108 (1988).

[4] B. Bollobás, P. Erdős, Graphs of extremal weights, *Ars Combin.* 50 (1998).

[5] P. Erdős, M. Saks, V. T. Sós, Maximum induced trees in graphs, *J. Combin. Theory Ser. B* 41 (1986).

[6] Y. Yang, L. Lu, The Randić index and the diameter of graphs, *Discrete Math.* 311 (2011), arXiv:1104.0426.

[7] W. Haemers, A. Mohammadian, B. Tayfeh-Rezaie, On the sum of Laplacian eigenvalues of graphs, *Linear Algebra Appl.* 432 (2010).

[8] J. H. Smith, Some properties of the spectrum of a graph, in *Combinatorial Structures and their Applications*, Gordon and Breach (1970).

[9] H. Deng, Z. Tang, J. Zhang, On a conjecture of Randić index and graph radius, *Filomat* (2015), arXiv:1210.2543.