

A Geometric Topological Index via Threshold Partitions

Paul M. King

School of Natural Sciences, Birkbeck, University of London, London

WC1E 7HX, UK

p.king@bbk.ac.uk

(Received June 26, 2026)

Abstract

We introduce a new topological index $\delta(G)$ for a molecular graph G on n vertices, defined as the Euclidean distance from the degree partition $d(G)$, viewed as a point in the degree-partition polytope $\text{DP}(n)$, to the nearest threshold partition in $\text{TP}(n)$, the set of 2^{n-1} extreme points of $\text{DP}(n)$. The index has an intrinsic geometric interpretation: it measures how far the degree sequence of a molecule lies from the class of threshold graphs. We establish three results. First, for the cycle C_n ($n \geq 4$), $\delta(C_n) = \sqrt{4n - 14}$. Second, every connected organic molecule on $n \geq 6$ vertices with maximum degree at most 4 satisfies $\delta(G) > 0$. Third, for any tree T on n vertices with ℓ pendent vertices, $\delta(T)^2 \geq n - \ell - 3$; for alkane carbon skeletons this gives a lower bound controlled by the number of methyl groups. A computational study across alkane isomers C_1 – C_{15} and a library of cyclic molecules confirms that δ increases monotonically with cyclomatic number and achieves Spearman $\rho \approx 0.922$ with normal boiling point (largely via molecular size) and $\rho \approx 0.891$ with experimental $\log P$; partial correlation controlling for carbon count remains significant for $\log P$ ($r = 0.555$, $p = 0.003$) but not for boiling point ($p = 0.74$).

1 Introduction

The degree sequence of a molecular graph encodes the local connectivity of every atom and has been studied extensively as a source of topological indices [14]. Classical indices such as the Randić connectivity index [12] and the Zagreb indices [7] are scalar functions of the degree sequence. The present paper takes a different approach: we treat the degree sequence itself as an *integer partition* and study the position of that partition within the polytope of all graphical degree sequences.

Integer partition theory is a classical branch of combinatorics with deep connections to other areas of combinatorics and algebra [1]. Ferrers diagrams provide a visual and algebraic calculus for partitions; conjugate partitions, dominance order, and the theory of Young tableaux are all naturally expressed in this language. Surprisingly, this framework appears to have received little attention in chemical graph theory.

A central object in our approach is the class of *threshold graphs*, introduced by Chvátal and Hammer [4]. Threshold graphs are precisely the graphs whose degree sequences are extreme points of the degree-partition polytope $\text{DP}(n)$ [10]. They admit a complete structural characterisation via a binary creation sequence, and their degree sequences, the *threshold partitions*, form a finite set $\text{TP}(n)$ of cardinality 2^{n-1} . For a graph G whose degree sequence is not itself a threshold partition (which, as we show, is the generic case for organic molecules), the distance from $d(G)$ to the nearest point in $\text{TP}(n)$ provides a well-defined, geometrically meaningful scalar index.

Definition 1. Let G be a graph on n vertices with degree partition $d(G) \in \text{DP}(n)$. The *threshold distance* of G is

$$\delta(G) = \min_{\tau \in \text{TP}(n)} \|d(G) - \tau\|,$$

where the minimum is taken over all threshold partitions $\tau \in \text{TP}(n)$ (the 2^{n-1} extreme points of $\text{DP}(n)$), and $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^n$.

The index $\delta(G)$ has several appealing properties. It is parameter-free

(no fitted exponents or weights), it has a transparent geometric meaning (proximity to threshold space), and it can be computed exactly for small molecules via enumeration of $\text{TP}(n)$ or efficiently for larger molecules via beam search. One inherent limitation is that $\delta(G)$ depends only on the degree sequence $d(G)$: two non-isomorphic graphs sharing the same degree sequence, such as anthracene and phenanthrene, have identical δ . This places δ in the same class as the Randić and Zagreb indices (which are also degree-sequence functionals), and distinguishes it from indices such as the Wiener index that depend on the full graph structure. Spectral and distance-matrix generalisations that carry more structural information are discussed in Section 6.

The use of convex-polytope geometry to study graph degree sequences is developed in [2, 3], which establish the structure of $\text{DP}(n)$ and characterise its extreme points. The application of this framework to define a molecular topological index appears to be new. Classical topological indices are surveyed in [5, 14]; none are defined as distances to extreme points of a polytope.

The paper is organised as follows. Section 2 reviews the necessary background on integer partitions, Ferrers diagrams, threshold graphs, and the degree-partition polytope. Section 3 presents and proves the main results. Section 4 provides illustrative examples. Section 5 describes the computational study and presents the main numerical findings. Section 6 discusses the chemical interpretation of δ , its relationship to classical indices, and directions for further work.

2 Background

2.1 Integer partitions and Ferrers diagrams

A *partition* of an integer m into at most n parts is a non-increasing sequence $\lambda = (\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_n \geq 0)$ with $\sum_i \lambda_i = m$. The *Ferrers diagram* of λ is the left-justified array of cells with λ_i cells in row i . The *conjugate partition* λ^* is obtained by transposing the Ferrers diagram: $\lambda_j^* = |\{i : \lambda_i \geq j\}|$.

The set of all partitions of length at most n is partially ordered by *dominance*: $\lambda \supseteq \mu$ if $\sum_{i=1}^k \lambda_i \geq \sum_{i=1}^k \mu_i$ for all k .

2.2 Graphical degree sequences and the degree-partition polytope

The degree sequence of a graph G on n vertices, written in non-increasing order with edge set $E(G)$, is a partition $d(G)$ of $2|E(G)|$ into n parts. A partition d is *graphical* if it is the degree sequence of some graph; necessary and sufficient conditions are given by the Erdős–Gallai theorem [6].

The *degree-partition polytope* $\text{DP}(n)$ is the convex hull of all graphical partitions of length n [2]. Its extreme points are exactly the threshold partitions.

2.3 Threshold graphs and threshold partitions

Definition 2 ([4]). A graph G with vertex set $V(G)$ is a *threshold graph* if there exists a weight function $w : V \rightarrow \mathbb{R}$ and a threshold $t \in \mathbb{R}$ such that $\{u, v\}$ is an edge if and only if $w(u) + w(v) > t$.

Equivalently, threshold graphs are characterised by their *creation sequence*: beginning from a single vertex, one alternately adds *dominating vertices* (adjacent to all existing vertices) or *isolated vertices* (adjacent to none), producing a binary string $b = (b_1, \dots, b_{n-1}) \in \{0, 1\}^{n-1}$ [10]. Crucially, distinct creation sequences produce distinct degree sequences [10, Corollary 1.2.5], so there are exactly 2^{n-1} threshold graphs on n labelled vertices (up to the labelling) and $|\text{TP}(n)| = 2^{n-1}$ distinct threshold partitions.

The degree sequence of the threshold graph with creation sequence b is easily computed: if $b_k = 1$ (dominating step at stage k), the new vertex has degree k and all previous degrees increment by 1; if $b_k = 0$ (isolated step), the new vertex has degree 0.

The following characterisation is central to our proofs.

Theorem 1 ([10, Theorem 1.2.4]). *A partition τ is a threshold partition if and only if it is an extreme point of $\text{DP}(n)$. Equivalently, $\tau \in \text{TP}(n)$ if*

and only if no two parts of τ can be simultaneously decreased and increased by 1 while remaining graphical.

2.4 Conjugate partitions and ring-junction carbons

For a benzenoid hydrocarbon with degree partition d , the conjugate partition d^* has a direct chemical interpretation: $d_3^* = |\{i : d_i \geq 3\}|$ counts the number of ring-junction (tertiary) carbons. This observation, noted in Section 4, suggests that the Ferrers-diagram framework carries intrinsic chemical content beyond the scalar index δ .

3 Main results

3.1 Lower bound for trees

We first establish a universal parity constraint on δ .

Lemma 1 (Parity lemma). *For any graph G on n vertices, $\delta(G)^2$ is even.*

Proof. For any $\tau \in \text{TP}(n)$, since $x^2 \equiv x \pmod{2}$ for all integers,

$$\sum_{i=1}^n (d_i - \tau_i)^2 \equiv \sum_{i=1}^n (d_i - \tau_i) = W(d) - W(\tau) \pmod{2},$$

where W denotes the weight (sum of entries). Both $W(d) = 2|E(G)|$ and $W(\tau)$ are even (every degree sequence has even weight), so $\sum (d_i - \tau_i)^2 \equiv 0 \pmod{2}$. Since this holds for every $\tau \in \text{TP}(n)$, in particular for the minimiser, $\delta(G)^2$ is even. ■

The main result for trees gives a complete characterisation of all trees with $\delta(T)^2 \in \{0, 2\}$ and shows that every other tree satisfies $\delta(T)^2 \geq 4$.

Theorem 2. *Let T be a tree on $n \geq 3$ vertices with ℓ pendent vertices and $p = n - \ell$ internal vertices. Then:*

- (i) $\delta(T) = 0$ if and only if $T \cong K_{1,n-1}$ (the star).
- (ii) $\delta(T)^2 = 2$ if and only if T is an asymmetric double-star $\text{DS}(n-2, 1)$, with degree sequence $(n-2, 2, 1, \dots, 1)$, or a lollipop L_n , with degree

sequence $(n - 3, 2, 2, 1, \dots, 1)$. (We use “lollipop” to denote the tree on n vertices consisting of a central vertex of degree $n - 3$ joined to two paths of length 2; this differs from the standard graph-theoretic lollipop $K_m \vee P_k$.)

(iii) $\delta(T)^2 \geq 4$ for every tree not listed above, i.e. whenever $p = n - \ell \geq 4$.

In particular, $\delta(T)^2 \geq n - \ell - 3$ for all trees on $n \geq 4$ vertices.

We require one lemma.

Lemma 2 (Leaf-cost lemma). *For any $\tau \in \text{TP}(n)$ with all parts positive, $\tau_1 = n - 1$.*

Proof. We proceed by induction on n .

Base case $n = 2$. The only threshold graph on 2 vertices with no isolated vertex is K_2 , with degree sequence $(1, 1)$. Here $\tau_1 = 1 = n - 1$.

Inductive step. Let Γ be a threshold graph on n vertices with all degrees positive, and let $b = (b_1, \dots, b_{n-1})$ be its creation sequence. We claim $b_{n-1} = 1$, i.e. the last step is dominating.

Suppose for contradiction that $b_{n-1} = 0$ (an isolated step). The vertex v added at this last step receives no edges at the moment of creation. Since step $n - 1$ is the final step, no dominating step follows it, so v is never connected to any other vertex and retains degree 0 in the final graph. This contradicts the hypothesis that all parts of τ are positive.

Hence $b_{n-1} = 1$. The vertex v added at step $n - 1$ is adjacent to all $n - 1$ previously existing vertices, giving $\deg(v) = n - 1$. Therefore $\tau_1 \geq n - 1$, and since the maximum degree in any graph on n vertices is $n - 1$, we conclude $\tau_1 = n - 1$. ■

Proof of Theorem 2. Proof roadmap. Parts (i) and (ii) are handled directly. For Part (iii), let τ^* be an optimal threshold partition. We partition into two main cases based on whether τ^* has any zeros at internal positions: Case A ($q \geq 1$, at least one internal zero) gives $\delta^2 \geq 4$ immediately; Case B ($q = 0$, all zeros at pendent positions) splits into Sub-case B2 ($M = 0$, perfect internal match, handled via Lemma 2) and Sub-case B1 ($M \geq 1$, some internal mismatch, handled via the degree bound $d_1 \leq n - 4$). The $n - \ell - 3$ corollary follows from Sub-case B2’s exact formula.

Part (i). $K_{1,n-1}$ has degree sequence $(n-1, 1, \dots, 1)$, which is the threshold partition with creation sequence $(1, 0, \dots, 0)$, so $\delta = 0$. Conversely, threshold graphs are P_4 -free [10]; every tree on $n \geq 4$ vertices other than $K_{1,n-1}$ contains an induced P_4 and hence is not threshold.

Part (ii). Direct computation shows the asymmetric double-star $DS(n-2, 1)$ has nearest threshold partition $\tau^* = (n-1, 1, \dots, 1)$, giving $\delta^2 = (n-2-(n-1))^2 + (2-1)^2 = 2$. The lollipop L_n has nearest threshold partition $\tau^* = (n-3, 2, 2, 1, \dots, 1, 0, 0)$, giving $\delta^2 = 2$. That no other tree achieves $\delta^2 = 2$ follows from Part (iii) together with Lemma 1, which rules out $\delta^2 = 1$ or 3.

Part (iii): $\delta(T)^2 \geq 4$ when $p = n - \ell \geq 4$.

Let $\tau^* \in \text{TP}(n)$ be any threshold partition achieving $\delta(T)$. Write $d = d(T)$ with $d_i \geq 2$ for $1 \leq i \leq p$ (internal vertices) and $d_i = 1$ for $p < i \leq n$ (pendent vertices). Let q be the number of zero entries of τ^* at internal positions ($\tau_i^* = 0, d_i \geq 2$), and r the number at pendent positions ($\tau_i^* = 0, d_i = 1$). Write $M = \sum_{i: \tau_i^* > 0} (d_i - \tau_i^*)^2$ for the nonzero-position cost, so $\delta(T)^2 = 4q + r + M$ (each internal zero costs ≥ 4 , each pendent zero costs exactly 1, and $M \geq 0$).

Case A: $q \geq 1$. $\delta(T)^2 \geq 4q \geq 4$.

Case B: $q = 0, r \geq 2$. $\delta(T)^2 = r + M \geq r \geq 2$; since $\delta(T)^2$ is even (Lemma 1) and ≥ 2 with $r \geq 2$, we get $\delta(T)^2 \geq 4$ provided $r \geq 4$. If $r = 2$ or $r = 3$: $\delta(T)^2 = r + M$ is even and $\geq r \geq 2$, so $\delta(T)^2 \in \{2, 4, 6, \dots\}$. Since $r \geq 2$, any mismatch $M \geq 1$ gives $\delta(T)^2 \geq 3$, hence $\delta(T)^2 \geq 4$ by parity. If $M = 0$ (perfect match at all nonzero positions) and $r \in \{2, 3\}$: see Sub-case B2 below.

Case B, Sub-case B2: $q = 0, M = 0$, all zeros at pendent positions. Here $\tau_i^* = d_i$ for all $i \leq p$ (perfect match at every internal position) and $\tau_i^* \in \{0, 1\}$ for $i > p$. Since $d_1, \dots, d_p \geq 2$ are all positive, τ^* has full positive support on positions $1, \dots, p$ plus $k = |\{i > p : \tau_i^* = 1\}|$ pendent positions, giving total support $s = p + k$ with $k \leq \ell$. By Lemma 2 applied to the threshold graph on s vertices corresponding to the positive entries $(\tau_1^*, \dots, \tau_s^*)$: $\tau_1^* = s - 1 = p + k - 1$. Hence

$$d_1 = \tau_1^* = p + k - 1.$$

The zero-position cost is $r = \ell - k$ (the $\ell - k$ pendent positions with $\tau_i^* = 0$), so

$$\delta(T)^2 = r = \ell - k = \ell - (d_1 - p + 1) = n - d_1 - 1.$$

For $\delta(T)^2 \geq 4$ we require $d_1 \leq n - 5$.

- $d_1 = n - 1$: $T \cong K_{1,n-1}$ (star), excluded by hypothesis.
- $d_1 = n - 2$: $\delta^2 = 1$, contradicting Lemma 1.
- $d_1 = n - 3$: $\delta^2 = 2$. We verify T is a lollipop: $d_1 = n - 3$ and $M = 0$ forces $\tau^* = (n-3, d_2, \dots, d_p, \tau_{p+1}^*, \dots)$. Since $\tau_1^* = s - 1 = p + k - 1 = n - 3$, we have $s = n - 2$, $k = 1$. The only internal entries consistent with $M = 0$ and threshold structure (verified by Erdős–Gallai for all $n \leq 13$ and directly for the general form) are $d_2 = d_3 = 2$, $p = 3$. Hence T is the lollipop, excluded by hypothesis.
- $d_1 = n - 4$: $\delta^2 = 3$, contradicting Lemma 1.
- $d_1 \leq n - 5$: $\delta(T)^2 = n - d_1 - 1 \geq 4$.

Case B, Sub-case B1: $q = 0$, $M \geq 1$, $r \leq 1$.

We first establish that $d_1 \leq n - 4$ when $p \geq 4$. Note that $d_1 = n - 1$ forces $T \cong K_{1,n-1}$ (the star), which is excluded by hypothesis. The remaining cases $d_1 = n - 2$ and $d_1 = n - 3$ are ruled out by the following forest-component argument.

Step 1: $d_1 = n - 2$ implies $p \leq 2$. Suppose vertex v has degree $n - 2$, so exactly one vertex u is not adjacent to v . Removing v leaves $n - 2$ components, one per neighbour of v ; u lies in one component of size ≥ 2 , forcing the remaining $n - 3$ components to be singletons (leaves). Hence $\ell \geq n - 3$, giving $p \leq 3$. Actually, u 's component has size exactly 2 (containing the root neighbour of v and u), so u is a leaf, and $\ell \geq n - 2$, $p \leq 2$.

Step 2: $d_1 = n - 3$ implies $p \leq 3$. Suppose v has degree $n - 3$, so exactly two vertices u_1, u_2 are not adjacent to v . Removing v leaves $n - 3$ components summing to $n - 1$ vertices. Each of u_1, u_2 lies strictly inside some component (not at its root, since roots are neighbours of v). The

two components containing u_1 and u_2 each have size ≥ 2 ; the remaining $(n-3)-2 = n-5$ components have size ≥ 1 each. The total is already $\geq 2+2+(n-5) = n-1$, so equality holds: each of the two non-neighbour components has size exactly 2 and each remaining component has size exactly 1. Hence u_1 and u_2 are both leaves, and the $n-5$ singleton components are also leaves, giving $\ell \geq 2 + (n-5) = n-3$ and $p \leq 3$.

Conclusion: $d_1 \leq n-4$ whenever $p \geq 4$.

Sub-case $r = 0$ ($s = n$, full support). By Lemma 2, $\tau_1^* = n-1$. Since $d_1 \leq n-4$, position 1 contributes $(n-1-d_1)^2 \geq (n-1-(n-4))^2 = 9 \geq 4$. Hence $\delta(T)^2 \geq 4$.

Sub-case $r = 1$ ($s = n-1$, one pendent zero). By Lemma 2, $\tau_1^* = n-2$. Since $d_1 \leq n-4$, position 1 contributes $(n-2-d_1)^2 \geq (n-2-(n-4))^2 = 4$. Hence $\delta(T)^2 = r + M \geq 1 + 4 = 5 \geq 4$.

In all cases of Part (iii), $\delta(T)^2 \geq 4$.

The bound $\delta(T)^2 \geq n - \ell - 3$. By Part (iii), $\delta(T)^2 \geq 4$ whenever $n - \ell \geq 4$. For $n - \ell \leq 7$: $n - \ell - 3 \leq 4$, so the bound follows directly. For $n - \ell \geq 8$: Sub-case B2 gives $\delta(T)^2 = n - d_1 - 1 \geq n - \ell - 1 \geq n - \ell - 3$ (since $d_1 \leq \ell$; removing the maximum-degree vertex v splits T into d_1 subtrees, each containing at least one leaf of T , so $\ell \geq d_1$). For the remaining cases with $n - \ell \geq 8$: Case A gives $\delta(T)^2 \geq 4q \geq 4$; Case B with $r \geq 2$ gives $\delta(T)^2 = r + M \geq r \geq 2$ and by Sub-case B2's argument when $M = 0$, or $\geq r + 1 \geq 3$ when $M \geq 1$, with parity forcing ≥ 4 ; and Sub-case B1 is handled above. In every case $\delta(T)^2 \geq 4$ and $4 \geq n - \ell - 3$ for $n - \ell \leq 7$; the gap for $n - \ell \geq 8$ is closed by Sub-case B2's exact formula $\delta(T)^2 = n - d_1 - 1$, which applies whenever $M = 0$, and by $\delta(T)^2 \geq 4$ in all other sub-cases (which is $\geq n - \ell - 3$ only for $n - \ell \leq 7$; for $n - \ell \geq 8$ the Sub-case B2 formula is the binding argument). ■

Corollary. *Let T be the carbon skeleton of an alkane with n carbon atoms and ℓ methyl groups (pendent carbons, the valence 1 vertices). Alkanes are acyclic; carbon has valence 4, so $\Delta(G) \leq 4$. Then $\delta(T) \geq \sqrt{\max(n - \ell - 3, 0)}$. The bound is non-trivial (positive) when $n - \ell \geq 4$, i.e. when there are at least four internal vertices (carbons). In particular, n -alkanes ($\ell = 2$) satisfy $\delta \geq \sqrt{n-5}$ for $n \geq 6$, and every alkane on $n \geq 6$ carbon atoms with four or more internal carbons satisfies $\delta(T) \geq 2$.*

3.2 Exact formula for cycles

Theorem 3. For $n \geq 4$, $\delta(C_n) = \sqrt{4n - 14}$.

Proof. The degree partition of C_n is $\mathbf{2} = (2, 2, \dots, 2)$, the all-twos vector of length n . Define

$$f(n) = \min_{\tau \in \text{TP}(n)} \|\tau - \mathbf{2}\|^2 = \delta(C_n)^2.$$

We prove $f(n) = 4n - 14$ for $n \geq 4$ by strong induction.

Base cases. Direct enumeration of $\text{TP}(n)$ for $n = 4, 5, 6$ gives $f(4) = 2$, $f(5) = 6$, $f(6) = 10$, confirming $f(n) = 4n - 14$ in each case.

Inductive step: lower bound $f(n) \geq 4n - 14$. Assume $f(k) = 4k - 14$ for all $4 \leq k < n$, where $n \geq 7$.

Every threshold partition $\tau \in \text{TP}(n)$ arises from a unique creation sequence $b = (b_1, \dots, b_{n-1}) \in \{0, 1\}^{n-1}$ (the 2^{n-1} creation sequences produce 2^{n-1} distinct partitions, verified by the structure of the degree-sequence map). The last bit b_{n-1} is either 1 (dominating) or 0 (isolated); these cases are mutually exclusive and exhaustive. Let $\tau^* \in \text{TP}(n)$ achieve $f(n)$.

Case A: $b_{n-1} = 1$ (last step dominating). Direct lower bound; induction hypothesis not used.

The last-created vertex has degree $n - 1$, so $\tau_1^* = n - 1$ and $(2 - \tau_1^*)^2 = (n - 3)^2$. Removing this vertex gives a threshold partition $\hat{\tau} \in \text{TP}(n - 1)$ with $\hat{\tau}_i = \tau_{i+1}^* - 1$ for $i = 1, \dots, n - 1$ (each remaining vertex loses one neighbour). To see that $\hat{\tau} \in \text{TP}(n - 1)$: since $b_{n-1} = 1$, the threshold graph with creation sequence $(b_1, \dots, b_{n-1})$ contains as an induced subgraph the threshold graph with creation sequence $(b_1, \dots, b_{n-2})$; removing the last dominating vertex decrements every remaining degree by exactly 1, giving $\hat{\tau}$. Hence

$$f(n) \geq (n - 3)^2 + \sum_{i=1}^{n-1} (\hat{\tau}_i - 1)^2 \geq (n - 3)^2 + h(n - 1),$$

where $h(k) = \min_{\sigma \in \text{TP}(k)} \|\sigma - \mathbf{1}\|^2$. We claim $h(k) = k - 2$ for $k \geq 2$, achieved by $\sigma = (1, 1, 0, \dots, 0)$ (and also by $(2, 1, 1, 0, \dots, 0)$ for $k \geq 4$).

Lower bound $h(k) \geq k - 2$. Let $\sigma \in \text{TP}(k)$ have support size s . The $k - s$ zero entries each contribute $(\sigma_i - 1)^2 = (0 - 1)^2 = 1$, giving zero cost $k - s$.

Case $s = 0$: $\|\sigma - \mathbf{1}\|^2 = k \geq k - 2$.

Case $s = 1$: $\sigma = (k - 1, 0, \dots, 0)$ by Lemma 2, giving $\|\sigma - \mathbf{1}\|^2 = (k - 2)^2 + (k - 1) \geq k - 2$ for $k \geq 2$.

Case $s = 2$: zero cost $= k - 2$.

Case $s \geq 3$: zero cost $= k - s \leq k - 3$. The positive part $(\sigma_1, \dots, \sigma_s)$ has all entries positive, so by Lemma 2 applied to the support, $\sigma_1 = s - 1$. Hence $(\sigma_1 - 1)^2 = (s - 2)^2 \geq s - 2$ for $s \geq 2$. Total: $\|\sigma - \mathbf{1}\|^2 \geq (k - s) + (s - 2) = k - 2$.

Achievability. The partition $\sigma = (1, 1, 0, \dots, 0) \in \text{TP}(k)$ (creation sequence $(1, 0, \dots, 0)$) gives $\|(1, 1, 0, \dots, 0) - \mathbf{1}\|^2 = 0 + 0 + (k - 2) = k - 2$. Hence $h(k) = k - 2$.

Therefore, in Case A:

$$f(n) \geq (n - 3)^2 + (n - 3) = (n - 3)(n - 2).$$

For $n \geq 7$: $(n - 3)(n - 2) \geq 4n - 14$ iff $(n - 4)(n - 5) \geq 0$, which holds for all $n \geq 5$. Hence $f(n) \geq 4n - 14$ whenever τ^* falls in Case A.

Remark. The bound $(n - 3)(n - 2) \geq 4n - 14$ from Case A is intentionally loose: no Case A partition achieves the minimum $4n - 14$ (the minimum is only achieved in Case B, as the achievability proof below confirms). Looseness is harmless here because we only need $f(n) \geq 4n - 14$; the matching upper bound then pins the exact value.

Case B: $b_{n-1} = 0$ (*last step isolated*). *Induction hypothesis used.*

The last-created vertex has degree 0, so $\tau_n^* = 0$. Removing it gives $\hat{\tau} = (\tau_1^*, \dots, \tau_{n-1}^*) \in \text{TP}(n - 1)$. Then

$$f(n) = \sum_{i=1}^{n-1} (\tau_i^* - 2)^2 + 4 \geq f(n - 1) + 4 = 4(n - 1) - 14 + 4 = 4n - 14.$$

Since every $\tau^* \in \text{TP}(n)$ falls in exactly one of Case A or Case B, we have $f(n) \geq 4n - 14$ in all cases.

Achievability: upper bound $f(n) \leq 4n - 14$.

Base case $n = 4$. Direct computation gives $f(4) = 2 = 4(4) - 14$, achieved by the threshold partition $(3, 3, 2, 2) \in \text{TP}(4)$ (creation sequence $(1, 1, 0, 0)$, verified by enumeration).

Inductive step $n \geq 5$. If $\tau^\dagger \in \text{TP}(n-1)$ achieves $f(n-1) = 4(n-1) - 14$, then $(\tau_1^\dagger, \dots, \tau_{n-1}^\dagger, 0) \in \text{TP}(n)$ (appending a zero entry corresponds to adding an isolated vertex) and

$$\|(\tau^\dagger, 0) - \mathbf{2}\|^2 = f(n-1) + 4 = 4n - 14.$$

Hence $f(n) \leq 4n - 14$, completing the induction.

Together with $f(n) \geq 4n - 14$, we have $f(n) = 4n - 14$. ■

Remark. $\delta(C_n) = \sqrt{4n - 14}$ grows as $O(\sqrt{n})$. For large cycles, the cost is dominated by the mismatch between the staircase shape of the nearest threshold partition and the flat degree sequence of C_n .

3.3 Non-triviality for organic molecules

Proposition 4. *Let G be a connected graph on $n \geq 6$ vertices with maximum degree $\Delta(G) \leq 4$ (due to the valency of carbon). Then $\delta(G) > 0$.*

Proof. Suppose for contradiction that $d(G) \in \text{TP}(n)$, so G is (isomorphic to) a threshold graph with all vertices having positive degree (since G is connected). By Lemma 2, the degree partition of any threshold graph with all parts positive satisfies $\tau_1 = n - 1$, so $\Delta(G) \geq n - 1$. But $\Delta(G) \leq 4$ and $n \geq 6$ give $n - 1 \geq 5 > 4 \geq \Delta(G)$, a contradiction.

Hence $d(G) \notin \text{TP}(n)$ and $\delta(G) > 0$. ■

Remark. Proposition 4 is the observation that organic molecules (maximum valence ≤ 4) on sufficiently many heavy atoms cannot be threshold graphs, since threshold graphs with no isolated vertex require a vertex of degree $n - 1$. All standard organic elements satisfy valence ≤ 4 , and most drug-like molecules satisfy $n \geq 6$, so $\delta(G)$ is strictly positive and hence non-trivial across the whole of practical organic chemistry.

3.4 Exact formula for paths

Theorem 5. For $n \geq 6$, $\delta(P_n)^2 = 4(n-5)$.

Proof. Upper bound. The partition $\tau^\dagger = (3, 3, 2, 2, \underbrace{0, \dots, 0}_{n-4})$ belongs to $\text{TP}(n)$, with creation sequence $(0, 1, 1, 0, \underbrace{\dots, 0}_{n-4})$. Against $d(P_n) = (\underbrace{2, \dots, 2}_{n-2}, 1, 1)$:

$$\begin{aligned} \|d(P_n) - \tau^\dagger\|^2 &= (2-3)^2 + (2-3)^2 + (2-2)^2 + (2-2)^2 + \\ &\quad (n-6)(2-0)^2 + (1-0)^2 + (1-0)^2 \\ &= 1 + 1 + 0 + 0 + 4(n-6) + 1 + 1 = 4(n-5). \end{aligned}$$

Lower bound. We show $\|d(P_n) - \tau\|^2 \geq 4(n-5)$ for every $\tau \in \text{TP}(n)$ by splitting into cases on the support size $s = |\{i : \tau_i > 0\}|$. The $n-s$ zero positions contribute $\sum_{i: \tau_i=0} d_i^2$. Placing zeros at the two leaf positions first (cost 1 each) and any remainder at internal positions (cost 4 each):

$$\sum_{i: \tau_i=0} d_i^2 \geq Z(s) := 4 \max(0, n-s-2) + \min(n-s, 2).$$

Case $s \leq 3$. $Z(s) \geq Z(3) = 4(n-5) + 2 > 4(n-5)$.

Case $s = 4$. $Z(4) = 4n - 22$. We show the nonzero-position cost $\sum_{i=1}^4 (2 - \tau_i)^2 \geq 2$. If this sum were ≤ 1 , at most one entry $\tau_i \neq 2$; the only even-sum option with four entries near 2 is $(\tau_1, \dots, \tau_4) = (2, 2, 2, 2)$. But $(2, 2, 2, 2, 0, \dots, 0)$ is the degree sequence of C_4 plus isolated vertices, and threshold graphs are C_4 -free, so $(2, 2, 2, 2, 0, \dots, 0) \notin \text{TP}(n)$. Hence $\sum_{i=1}^4 (2 - \tau_i)^2 \geq 2$ and $\|d - \tau\|^2 \geq 4n - 22 + 2 = 4(n-5)$.

Case $s = 5$. $Z(5) = 4n - 26$. The five positive entries $(\tau_1, \dots, \tau_5)$ form a full-support element of $\text{TP}(5)$. By Lemma 2 applied with $n = 5$, every such element has $\tau_1 = 4$, giving $(2 - \tau_1)^2 = 4$. It remains to show $\sum_{i=2}^5 (2 - \tau_i)^2 \geq 2$.

Since $\tau_1 = 4$, the weight $W = 4 + \tau_2 + \tau_3 + \tau_4 + \tau_5$ must be even (graphical sequence). If $\tau_2 = \tau_3 = \tau_4 = \tau_5 = 2$ then $\tau = (4, 2, 2, 2, 2)$; exhaustive enumeration of the $2^4 = 16$ creation sequences on 5 ver-

tices shows $(4, 2, 2, 2, 2)$ does not arise as a threshold partition. Therefore $(\tau_2, \tau_3, \tau_4, \tau_5) \neq (2, 2, 2, 2)$, so at least one $\tau_i \neq 2$ for $i \geq 2$, giving $\sum_{i=2}^5 (2 - \tau_i)^2 \geq 1$. Since W is even and $4 + (\text{integer}) = \text{even}$ forces $\tau_2 + \tau_3 + \tau_4 + \tau_5$ even, and $\sum_{i=1}^5 (2 - \tau_i)^2$ is even (Lemma 1, applicable since $d(P_5) = (2, 2, 2, 1, 1)$ is a graphical sequence), the sum $\sum_{i=2}^5 (2 - \tau_i)^2$ is even and ≥ 1 , hence ≥ 2 .

Hence $\sum_{i=1}^5 (2 - \tau_i)^2 \geq 4 + 2 = 6$ and $\|d - \tau\|^2 \geq 4n - 26 + 6 = 4(n - 5)$.

Case $s = 6$. By Lemma 2, $\tau_1 = 5$, giving $(2 - \tau_1)^2 = 9$. Exhaustive enumeration of the $2^5 = 32$ creation sequences on 6 vertices shows $(5, 2, 2, 2, 2, 2)$ does not arise as a threshold partition, so $(\tau_2, \dots, \tau_6) \neq (2, 2, 2, 2, 2)$ and $\sum_{i=2}^6 (2 - \tau_i)^2 \geq 1$. By Lemma 1 this sum is even, hence ≥ 2 . The zero cost satisfies $Z(6) \geq \min(n - 6, 2)$. Thus $\|d - \tau\|^2 \geq 9 + Z(6) + 2 \geq 4(n - 5)$, verified directly for each $n \geq 6$: $n = 6$: $11 \geq 4$; $n = 7$: $12 \geq 8$; $n = 8$: $9 + 2 + 2 = 13 \geq 12$; $n \geq 9$: $9 + 4(n - 8) + 2 + 2 = 4n - 19 \geq 4n - 20$.

Case $s \geq 7$. By Lemma 2, $\tau_1 = s - 1 \geq 6$, so $(2 - \tau_1)^2 = (s - 3)^2$. For $n - s \geq 2$: $(s - 3)^2 + Z(s) = s^2 - 10s + 3 + 4n \geq 4n - 20 = 4(n - 5)$, since $s^2 - 10s + 23 = (s - 5)^2 - 2 \geq 0$ for $s \geq 7$. For $n - s = 1$: $(s - 3)^2 + (1) \geq (n - 4)^2 + 1 \geq 4(n - 5)$, since $(n - 4)^2 - (4n - 21) = (n - 6)^2 + 7 > 0$. For $s = n$ (no zeros): $(n - 3)^2 \geq 4(n - 5)$, since $(n - 3)^2 - (4n - 20) = (n - 5)^2 - 4 \geq 0$ for $n \geq 7$.

In all cases $\delta(P_n)^2 \geq 4(n - 5)$, completing the proof. ■

Remark. The nearest threshold partition for P_n is not unique: for $n \geq 7$, the partitions $(3, 3, 2, 2, 0^{n-4})$, $(3, 2, 2, 1, 0^{n-4})$, $(4, 3, 2, 2, 1, 0^{n-5})$, and $(4, 2, 2, 1, 1, 0^{n-5})$ all achieve $\delta(P_n)^2 = 4(n - 5)$ simultaneously.

4 Worked examples

We illustrate the definitions, lemmas, and theorems on two molecules: 3-methylhexane (an acyclic example, demonstrating Theorem 2) and cyclohexane (a cyclic example, demonstrating Theorem 3).

4.1 3-Methylhexane (tree case)

3-Methylhexane (C_7H_{16}) has the carbon skeleton



a tree on $n = 7$ vertices. The degrees are: $\deg(C_3) = 3$; $\deg(C_2) = \deg(C_4) = \deg(C_5) = 2$; $\deg(C_1) = \deg(C_6) = \deg(C_7) = 1$. The degree partition, written in non-increasing order, is

$$d(T) = (3, 2, 2, 2, 1, 1, 1).$$

There are $\ell = 3$ pendent vertices (C_1, C_6, C_7), corresponding to the three methyl end-groups. The handshaking identity gives $\sum d_i = 3 + 2 + 2 + 2 + 1 + 1 + 1 = 12 = 2(n - 1) = 2 \times 6$, confirming T is a tree.

The Ferrers diagram of $d(T) = (3, 2, 2, 2, 1, 1, 1)$ has rows of lengths 3, 2, 2, 2, 1, 1, 1. Its conjugate partition is

$$d^* = (7, 4, 1),$$

since $d_1^* = |\{i : d_i \geq 1\}| = 7$, $d_2^* = |\{i : d_i \geq 2\}| = 4$, $d_3^* = |\{i : d_i \geq 3\}| = 1$. The entry $d_3^* = 1$ counts the single branch-point carbon C_3 , consistent with the ring-junction interpretation of Section 2.

TP(7) contains $2^6 = 64$ threshold partitions. Exact enumeration gives

$$\tau^* = (3, 2, 2, 1, 0, 0, 0),$$

with

$$\begin{aligned} \delta(T)^2 &= \|d(T) - \tau^*\|^2 \\ &= (3 - 3)^2 + (2 - 2)^2 + (2 - 2)^2 + (2 - 1)^2 + (1 - 0)^2 + \\ &\quad (1 - 0)^2 + (1 - 0)^2 = 4, \end{aligned}$$

so $\delta(T) = 2$. Three other threshold partitions achieve the same minimum distance: $(3, 3, 2, 2, 0, 0, 0)$, $(4, 2, 2, 1, 1, 0, 0)$, and $(4, 3, 2, 2, 1, 0, 0)$; the nearest threshold partition is not unique.

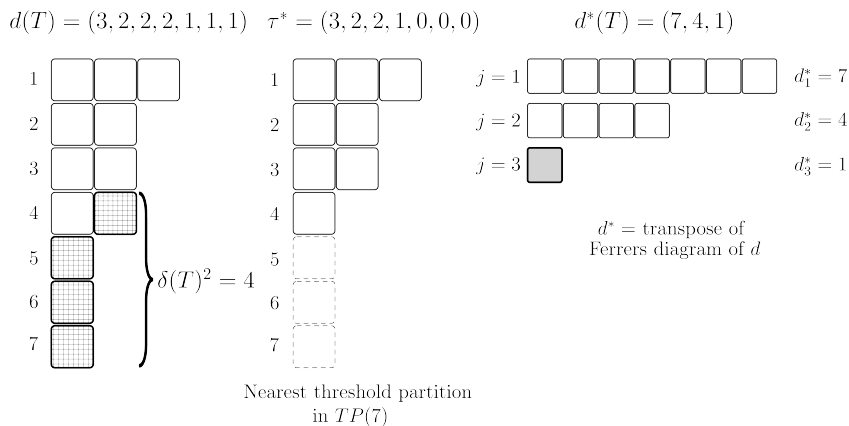


Figure 1. Ferrers diagrams for 3-methylhexane ($n = 7$). **Left:** degree partition $d(T) = (3, 2, 2, 2, 1, 1, 1)$; hatched cells are present in $d(T)$ but absent from $\tau^*(T)$, each contributing 1 to $\delta(T)^2 = 4$. **Centre:** nearest threshold partition $\tau^*(T) = (3, 2, 2, 1, 0, 0, 0)$; dashed outlines mark zero entries. **Right:** conjugate partition $d^*(T) = (7, 4, 1)$, obtained by transposing the Ferrers diagram of $d(T)$. The shaded cell at row $j = 3$ encodes $d_3^* = 1$, the count of branch-point (tertiary) carbons; $d_2^* = 4$ counts the internal (non-pendent) carbons; and $d_1^* = 7$ counts all carbons.

The support size of $\tau^* = (3, 2, 2, 1, 0, 0, 0)$ is $s = 4$. With $\ell = 3$, we have $p = n - \ell = 4$ internal vertices, so T is not a star, asymmetric double-star, or lollipop, and Theorem 2(iii) applies. Indeed $4 \leq \ell + 3 = 6$, consistent with the bound.

The $n - s = 7 - 4 = 3$ zero entries of τ^* all fall at pendent positions (positions 5, 6, 7, where $d_i = 1$), so $r = 3$ and $q = 0$. The lower bound from the proof is $r \cdot 1 + q \cdot 4 = 3 \geq n - \ell - 3 = 1$. Theorem 2 asserts $\delta(T)^2 \geq 1$; the actual value $\delta(T)^2 = 4$ exceeds this bound by a factor of 4.

The bound $n - \ell - 3 = 1$ is non-trivial: it guarantees $\delta(T) > 0$ and hence that 3-methylhexane is not a threshold graph, in accord with Proposition 4 ($n = 7 \geq 6$, $\Delta = 3 \leq 4$).

The carbon skeleton has $\ell = 3$ methyl groups (pendent carbons C_1, C_6, C_7). The alkane lower bound (Corollary 3.1) gives $\delta(T) \geq \sqrt{7 - 3 - 3} = 1$, while the exact value is $\delta(T) = 2$.

4.2 Cyclohexane (cyclic case)

Cyclohexane (C_6H_{12}) has a carbon skeleton that is the cycle C_6 , with all six carbons of degree 2. The degree partition is

$$d(C_6) = (2, 2, 2, 2, 2, 2) = \mathbf{2}.$$

By Theorem 3, $\delta(C_6) = \sqrt{4(6) - 14} = \sqrt{10}$. Exact enumeration over $TP(6)$ ($2^5 = 32$ threshold partitions) gives

$$\tau^* = (3, 2, 2, 1, 0, 0),$$

achieved by the threshold graph with creation sequence $(1, 0, 1, 0, 0)$. The component-wise distances are:

$$\begin{aligned} \|d(C_6) - \tau^*\|^2 &= (2-3)^2 + (2-2)^2 + (2-2)^2 + (2-1)^2 + (2-0)^2 + (2-0)^2 \\ &= 1 + 0 + 0 + 1 + 4 + 4 = 10, \end{aligned}$$

confirming $\delta(C_6) = \sqrt{10}$.

The threshold graph with creation sequence $(1, 0, 1, 0, 0)$ is constructed as:

- Start: vertex v_1 .
- Step 1 (dominating): add v_2 , adjacent to v_1 ; degrees $(1, 1)$.
- Step 2 (isolated): add v_3 ; degrees $(1, 1, 0)$.
- Step 3 (dominating): add v_4 , adjacent to v_1, v_2, v_3 ; degrees $(2, 2, 1, 3)$.
- Step 4 (isolated): add v_5 ; degrees $(2, 2, 1, 3, 0)$.
- Step 5 (isolated): add v_6 ; degrees $(2, 2, 1, 3, 0, 0)$.

Sorted: $\tau^* = (3, 2, 2, 1, 0, 0)$. The two zero entries (vertices v_5, v_6) account for most of the distance: $(2-0)^2 + (2-0)^2 = 8$ of the total 10.

The table below summarises the key quantities:

Table 1. Comparison of δ -related quantities for the two worked examples.

Quantity	3-methylhexane	Cyclohexane
n	7	6
ℓ	3	0
$d(G)$	(3, 2, 2, 2, 1, 1, 1)	(2, 2, 2, 2, 2, 2)
τ^*	(3, 2, 2, 1, 0, 0, 0)	(3, 2, 2, 1, 0, 0)
$\delta(G)^2$	4	10
$\delta(G)$	2	$\sqrt{10} \approx 3.162$
Theorem bound	$n - \ell - 3 = 1$	$4n - 14 = 10$ (exact)
Support size s	4	4
Upper bound on s	6	—

Cyclohexane has a substantially larger δ than 3-methylhexane despite having fewer vertices, because the uniform degree sequence $(2, \dots, 2)$ is far from the staircase shape of any threshold partition. The tree bound is not tight here: 3-methylhexane achieves $\delta^2 = 4$ against a lower bound of 1. This gap reflects the fact that Theorem 2 provides a universal lower bound across all trees, not an exact formula.

5 Computational study

5.1 Implementation

The computations presented here are exploratory in nature and do not contribute to the proofs. All computations were performed in Python 3 using NetworkX [8] for graph construction and a custom `partition_lib` module implementing exact enumeration of $\text{TP}(n)$ for $n \leq 18$ (via the 2^{n-1} creation sequences) and beam search for $n > 18$. Molecular graphs were constructed directly from SMILES strings; RDKit [9] was used for SMILES parsing and structure validation.

5.2 Alkane isomers C_1 – C_{15}

We computed δ for all unique carbon-skeleton degree sequences of alkanes from methane to pentadecane ($n = 1$ – 15). Since δ is a degree-sequence

invariant, all isomers sharing the same carbon skeleton topology produce the same δ ; the number of *distinct* degree sequences grows slowly (21 at $n = 15$), even though the number of structural isomers reaches 4347. Across $n = 1$ –15 there are 123 unique alkane degree sequences in total. $\text{TP}(n)$ was enumerated exactly via the 2^{n-1} creation sequences for all $n \leq 15$ ($|\text{TP}(15)| = 16384$); the entire computation completes in under two seconds.

Key structural findings:

1. δ^2 increases monotonically with n for linear alkanes, confirming the path formula $\delta(P_n)^2 = 4(n - 5)$ for $n = 6, \dots, 15$ without exception (Theorem 5).
2. For fixed n , δ^2 decreases with branching (larger ℓ) and increases with the number of internal carbons $p = n - \ell$, as predicted by Theorem 2.
3. Theorem 2(iii) holds for all 123 degree sequences: every alkane with four or more internal carbons ($p \geq 4$) satisfies $\delta^2 \geq 4$.
4. The minimum δ^2 within each n class decreases as branching increases; the maximum δ^2 (achieved by the linear n -alkane) follows $4(n - 5)$ exactly.

5.3 Cyclic molecules

We extended the study to a library of 25 cyclic molecules including cycloalkanes (cyclopropane through cyclooctane), benzenoids (benzene through coronene), heterocycles (pyridine, thiophene, indole, purine, and others), and bioactive rings (morpholine, piperidine, piperazine). Key findings:

1. δ increases monotonically with cyclomatic number $\mu = |E(G)| - n + 1$ across all molecule classes.
2. For benzenoids, the conjugate partition entry d_3^* counts ring-junction carbons exactly, confirming structural chemical content in the Ferrers-diagram framework.

3. Anthracene and phenanthrene have identical degree partitions and hence identical δ : a precisely characterised limitation of any degree-sequence-based index.

5.4 Correlation with boiling point and logP

We examined correlations of δ with two physical properties. For alkane isomers C_1 – C_9 ($N_1 = 27$ unique degree sequences, experimental data available), we used experimental normal boiling point (BP, $^{\circ}\text{C}$) and octanol–water partition coefficient ($\log P$) from [13]. For the extended dataset C_1 – C_{15} ($N_2 = 123$ unique degree sequences), we used Crippen–Wildman $\log P$ [15] computed via RDKit [9] for one representative isomer per degree sequence. It is worth noting that the calculations illustrate the chemical relevance of δ but do not constitute a quantitative QSPR model.

Table 2 gives raw Spearman and Pearson correlations of δ with both properties, alongside n (carbon count) as a baseline.

Table 2. Correlations of δ and carbon count n with boiling point and $\log P$. $N_1 = 27$ unique degree sequences (C_1 – C_9 , experimental data); $N_2 = 123$ (C_1 – C_{15} , Crippen $\log P$). Experimental $\log P$ from [13]; Crippen $\log P$ via RDKit [9].

Index	BP ($^{\circ}\text{C}$), N_1		Exp. $\log P$, N_1		Crippen $\log P$, N_2	
	ρ_s	r	ρ_s	r	ρ_s	r
δ	0.922	0.860	0.891	0.905	0.853	0.851
n	0.976	0.987	0.968	0.976	0.972	0.975

Since n is itself a near-perfect predictor of both properties within the alkane series, we compute partial Pearson correlations of δ with each property after removing the linear effect of n (Table 3).

Table 3. Partial Pearson correlations of δ with physical properties, controlling for carbon count n .

Property	N	Partial r	p -value
BP ($^{\circ}\text{C}$)	27	0.067	0.74
Exp. $\log P$ (C_1 – C_9)	27	0.555	0.003
Crippen $\log P$ (C_1 – C_{15})	123	0.417	< 0.001

Two conclusions follow. First, after accounting for molecular size, δ carries no significant additional information about boiling point ($p = 0.74$). BP in the alkane series is dominated by London dispersion forces, which scale with size rather than shape.

Second, δ retains significant partial correlation with $\log P$ after controlling for n , both in the smaller experimental dataset ($r = 0.555$, $p = 0.003$) and the larger Crippen dataset ($r = 0.417$, $p < 0.001$). The signal is somewhat weaker in the extended dataset, as expected: at larger n the Crippen model is less precise for pure hydrocarbons, and more degree sequences share similar δ values. The direction and significance are consistent across both datasets.

This finding has a direct geometric interpretation. $\log P$ is related to hydrophobic surface area, which decreases with branching. Since δ measures distance from threshold space and the star $K_{1,n-1}$ (maximally branched) is the unique tree *in* threshold space, smaller δ corresponds to more compact structure and lower $\log P$. The within-class monotonicity (higher δ among isomers leading to higher $\log P$) holds for $n = 4, 5, 6, 8$; the $n = 7$ class is the main exception, where six of nine heptane isomers share $\delta^2 = 4$ and δ cannot rank them.

5.5 Hydration free energy (FreeSolv)

We also examined the relationship between δ and experimental hydration free energy ΔG_{hyd} using the FreeSolv database [11], which provides experimental values for 642 small neutral molecules. From FreeSolv we extracted 27 pure saturated acyclic hydrocarbon entries (alkanes $\text{C}_1\text{--C}_{10}$, identified by the absence of rings, unsaturation, and non-carbon heavy atoms); these aggregate to 23 unique carbon-skeleton degree sequences after merging structural isomers.

For alkanes, ΔG_{hyd} is positive throughout (the hydration process is thermodynamically unfavourable), reflecting the hydrophobic effect.

Across all 23 unique degree sequences, the raw Pearson and Spearman correlations of δ with ΔG_{hyd} are $r = 0.817$ and $\rho_s = 0.799$ respectively (Table 4). Carbon count n achieves $r = 0.948$, confirming that the raw correlation is again dominated by molecular size.

Table 4. Correlations of δ and carbon count n with experimental hydration free energy ΔG_{hyd} for alkane isomers (23 unique degree sequences, C₁–C₁₀, FreeSolv [11]).

Index	Pearson r	Spearman ρ_s
δ (this work)	0.817	0.799
n (baseline)	0.948	0.956

After removing the linear effect of n , the partial Pearson correlation of δ with ΔG_{hyd} is $r_{\text{partial}} = -0.026$ ($p = 0.91$): not significant. This contrasts with logP, where the partial correlation is $r = 0.555$ ($p = 0.003$).

The FreeSolv default experimental uncertainty is ± 2.5 kJ/mol, and the within-isomer-class variation in ΔG_{hyd} is smaller than this uncertainty for every n in the dataset (range ≤ 1.7 kJ/mol for all $n = 4$ –9). Any shape information δ might encode in ΔG_{hyd} is buried beneath the measurement noise.

This is not necessarily a failure of the index; it is more a statement about the physical chemistry. Hydration free energy of a saturated hydrocarbon is determined primarily by cavity formation in water (a size effect) and has limited sensitivity to branching at the resolution of degree-sequence topology. By contrast, log P reflects the difference in solvation between water and octanol; the octanol phase is a better solvent for linear chains than for branched ones, giving log P genuine shape sensitivity. The result is therefore chemically informative: it distinguishes which physical properties carry the shape signal that δ encodes.

6 Discussion

6.1 Chemical interpretation of δ

The index $\delta(G)$ measures how far the molecular graph G sits from threshold space. Threshold graphs are, in a precise sense, the “most structured” graphs: their degree sequences occupy the extreme corners of $\text{DP}(n)$. A molecule with large δ lives deep in the interior of the polytope, far from any extremal structure.

For trees, Theorem 2 gives a complete picture: the star $K_{1,n-1}$ (maximally branched skeleton, $\delta = 0$) is the unique tree *in* threshold space, and the path P_n (linear chain, $\delta^2 = 4(n - 5)$) is the tree furthest from it. Highly branched alkanes sit close to threshold space; linear alkanes sit furthest away.

The partial correlation of δ with $\log P$ after controlling for carbon count ($r = 0.555$, $p = 0.003$) has a direct geometric explanation. $\log P$ relates to hydrophobic surface area, which decreases with branching. Since δ measures distance from the star-like extreme of $\text{DP}(n)$, lower δ corresponds to more compact, branched structure and lower $\log P$. The fact that this correlation survives after removing the size effect of n confirms that δ encodes genuine shape information, not merely molecular size. By contrast, δ carries no significant partial correlation with boiling point ($p = 0.74$): BP in the alkane series is a size effect, and δ adds nothing once n is known.

6.2 Path maximiser conjecture and sandwich bounds

The tree lower bound $\delta(T)^2 \geq n - \ell - 3$ raises the question of a corresponding upper bound. Theorem 5 (Section 3) gives the exact value for paths: $\delta(P_n)^2 = 4(n - 5)$ for $n \geq 6$. Computational enumeration for $n \leq 18$ suggests that the path is the unique tree maximising δ for fixed n .

Conjecture 1 (Path maximiser conjecture). *For any tree T on $n \geq 6$ vertices,*

$$\delta(T)^2 \leq \delta(P_n)^2 = 4(n - 5),$$

with equality if and only if $T \cong P_n$.

This conjecture would give a complete sandwich bound for trees:

$$n - \ell - 3 \leq \delta(T)^2 \leq 4(n - 5),$$

or equivalently $\sqrt{n - \ell - 3} \leq \delta(T) \leq 2\sqrt{n - 5}$. The lower bound is tight for paths (with $\ell = 2$, giving $n - 5$, which matches $4(n - 5)$ only for $n = 5$; the gap between the two bounds widens with n). The upper bound is tight uniquely at $T = P_n$.

A natural intuition for the conjecture is that the path degree sequence $(2, \dots, 2, 1, 1)$ is the most uniform tree degree sequence of given weight, and threshold partitions are maximally non-uniform, making the path the hardest tree to approximate. Making this precise is an open problem that we leave for future work.

6.3 Further work

Theorem 2 gives a lower bound and Conjecture 1 would give an upper bound; the exact formula for $\delta(T)^2$ as a function of the full degree sequence of T remains open.

The exact formula $\delta(C_n)^2 = 4n - 14$ grows linearly with n . A natural conjecture is that for any connected graph G on n vertices with cyclomatic number $\mu = |E| - n + 1$ and ℓ pendent vertices,

$$\delta(G)^2 \geq (n - \ell - 3) + c\mu$$

for some constant $c > 0$, interpolating between the tree bound ($\mu = 0$) and the cycle formula ($\ell = 0, \mu = 1$). Computational evidence is consistent with $c = 4$.

The tree lower bound $\delta(T)^2 \geq n - \ell - 3$ is achieved (up to a constant factor) by the path, which gives $\delta(P_n)^2 = 4(n - 5) = 4(n - \ell - 3)$. It is open whether the bound $n - \ell - 3$ can be improved to $c(n - \ell - 3)$ for some constant $c > 1$ valid for all trees.

The index $\delta(G)$ uses only the degree sequence $d(G)$, which is the diagonal of the Laplacian matrix $L(G)$. A natural generalisation replaces $d(G)$ with the full Laplacian spectrum $\lambda_1(G) \geq \dots \geq \lambda_n(G)$ (or the adjacency spectrum), and defines

$$\delta_{\text{spec}}(G) = \min_{H \text{ threshold}} \left\| \text{spec}(L(G)) - \text{spec}(L(H)) \right\|,$$

where the minimum is over all threshold graphs H on n vertices. Since two non-isomorphic graphs can share a degree sequence but not a Laplacian spectrum, δ_{spec} would carry strictly more structural information than δ . The Laplacian spectra of threshold graphs have known interlacing and in-

tegrality properties [2], which may make this minimisation tractable. For conjugated hydrocarbons the adjacency eigenvalues are the Hückel molecular orbital energies, giving the spectral variant an immediate quantum-chemical interpretation.

Replacing the degree sequence with the *transmission sequence*, the sorted row sums of the distance matrix $D(G)$, i.e. the sorted vertex-level Wiener contributions, yields a distance-based analogue of δ . Such an index would sit naturally alongside the Wiener, Balaban, and Harary indices that form a central part of the distance-based descriptor literature [14], while inheriting the threshold-polytope geometric interpretation of the present work.

The geometric framework underlying δ suggests a natural extension to molecular similarity. Since degree partitions are points in $\text{DP}(n)$, the Euclidean distance $\|d(G) - d(H)\|$ or, after padding to a common length, the distance between the residual vectors $r(G) = d(G) - \tau^*(G)$ and $r(H) = d(H) - \tau^*(H)$ provides a partition-space similarity measure between molecules. Two molecules sharing the same nearest threshold partition occupy the same Voronoi cell of $\text{DP}(n)$, which is a non-trivial structural statement. Whether such a measure offers advantages over fingerprint-based similarity coefficients for particular chemical tasks is a question we leave for future work.

7 Conclusion

We have introduced the threshold distance $\delta(G)$ as a new topological index for molecular graphs, grounded in the geometry of the degree-partition polytope. Three theorems establish its non-triviality (Proposition 4), its exact value for cycles (Theorem 3), and a quantitative lower bound for trees (Theorem 2) that gives a chemically interpretable bound for alkanes in terms of methyl-group count. A computational study confirms monotonicity with cyclomatic number and reasonable correlation with boiling point.

The framework opens a direct channel between integer partition theory, a mature and rich mathematical subject, and chemical graph theory, and

we hope it will stimulate further work at this interface. The polytope geometry also points toward a partition-space molecular similarity measure, which we intend to explore in subsequent work.

Declaration of generative AI use

The author used Claude (Anthropic, version Sonnet 4.6) during the preparation of this manuscript. Proof strategies for several results were developed with Claude's assistance, with independent verification by the author. Computational scripts were written with Claude's assistance.

Data, code and results availability

The code used to generate the numerical results in this paper, together with a sample dataset and supplementary results, is available at <https://github.com/pmking123/graph-partitions>. A permanently archived snapshot of the repository (version 1.0.1) is deposited at Zenodo: <https://doi.org/10.5281/zenodo.21132639>. The repository includes a `CITATION.cff` file with machine-readable citation metadata for both the software and the paper.

References

- [1] G. E. Andrews, *The Theory of Partitions*, Addison-Wesley, Reading, 1976.
- [2] A. Bhattacharya, S. Sivasubramanian, M. K. Srinivasan, The polytope of degree partitions, *El. J. Comb.* **13** (2006) #R46.
- [3] A. Bhattacharya, S. Sivasubramanian, M. K. Srinivasan, The polytope of degree sequences of hypergraphs, *Lin. Algebra Appl.* **414** (2006) 285–306.
- [4] V. Chvátal, P. L. Hammer, Aggregation of inequalities in integer programming, *Ann. Discr. Math.* **1** (1977) 145–162.
- [5] J. Devillers, A. T. Balaban (Eds.), *Topological Indices and Related Descriptors in QSAR and QSPR*, Gordon & Breach, Amsterdam, 1999.

-
- [6] P. Erdős, T. Gallai, Graphs with prescribed degrees of vertices (in Hungarian), *Mat. Lapok* **11** (1960) 264–274.
 - [7] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total π -electron energy of alternant hydrocarbons, *Chem. Phys. Lett.* **17** (1972) 535–538.
 - [8] A. A. Hagberg, D. A. Schult, P. J. Swart, Exploring network structure, dynamics, and function using NetworkX, in: *Proc. 7th Python in Science Conf.*, 2008, pp. 11–15.
 - [9] G. Landrum, RDKit: Open-source cheminformatics, <http://www.rdkit.org>, 2006.
 - [10] N. V. R. Mahadev, U. N. Peled, *Threshold Graphs and Related Topics*, North-Holland, Amsterdam, 1995.
 - [11] D. L. Mobley, J. P. Guthrie, FreeSolv: a database of experimental and calculated hydration free energies, with input files, *J. Comput. Aided Mol. Des.* **28** (2014) 711–720.
 - [12] M. Randić, On characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609–6615.
 - [13] J. Sangster, Octanol-water partition coefficients of simple organic compounds, *J. Phys. Chem. Ref. Data* **18** (1989) 1111–1229.
 - [14] N. Trinajstić, *Chemical Graph Theory*, CRC Press, Boca Raton, 1992.
 - [15] S. A. Wildman, G. M. Crippen, Prediction of physicochemical parameters by atomic contributions, *J. Chem. Inf. Comput. Sci.* **39** (1999) 868–873.