Scientific Reports

https://doi.org/10.1038/s41598-026-53478-4

Article in Press

Prodeg type Shannon graph entropies with closed forms bounds and QSPR modeling

Received: 24 January 2026

Accepted: 12 May 2026

Published online: 20 May 2026

Cite this article as: Alsharafi M. & Zeren Y. Prodeg type Shannon graph entropies with closed forms bounds and QSPR modeling. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-53478-4

Mohammed Alsharafi & Yusuf Zeren

We are providing an unedited version of this manuscript to give early access to its findings. Before final publication, the manuscript will undergo further editing. Please note there may be errors present which affect the content, and all legal disclaimers apply.

If this paper is publishing under a Transparent Peer Review model then Peer Review reports will publish with the final article.

ARTICLE IN PRESS

© The Author(s) 2026. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

ARTICLE IN PRESS

Prodeg Type Shannon Graph Entropies with Closed Forms Bounds and QSPR Modeling

Mohammed Alsharafi [1,2,*], Yusuf Zeren[2]

[1]Department of Mechatronics Engineering, Faculty of Engineering and Computing, University of Science and Technology, Yemen

[2]Department of Mathematics, Faculty of Arts and Science, Yildiz Technical University, 34220 Esenler, Istanbul, Turkey

alsharafi205010@gmail.com/ m.s.alsharafi@ust.edu (M.A.), yzeren@yildiz.edu.tr(Y.Z.)

Abstract

Degree-based graph entropies quantify structural heterogeneity by transforming vertex-degree information into a probability distribution and applying Shannon entropy. We develop a unified framework for three Prodeg-type degree-power invariants, namely the Inverse Prodeg index $IP(G) = \sum_{v\in V(G)} d_G(v)^{1/2}$, the Misbalance Prodeg index $MP(G) = \sum_{v\in V(G)} d_G(v)^{3/2}$, and the Yemen Prodeg index $YP(G) = \sum_{v\in V(G)} d_G(v)^{5/2}$, together with their associated entropies $E_{IP}(G)$, $E_{MP}(G)$, and $E_{YP}(G)$. More generally, for $a > 0$ we consider the degree-weighted distribution $p_v \propto d_G(v)^a$ and the Shannon entropy $E_a(G)$, recovering the Prodeg cases at $a \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\}$. We derive closed-form expressions for representative graph families (complete graphs, cycles, paths, stars, and complete bipartite graphs) and establish sharp extremal behavior: for connected graphs on $n \geqslant 2$ vertices, $E_{IP}(G), E_{MP}(G), E_{YP}(G) \leqslant \log n$, with equality if and only if G is regular, while highly imbalanced families (e.g., stars) exhibit strong concentration and vanishing entropy for $a > 1$ as $n \to \infty$. We further provide explicit two-sided bounds in terms of degree extremes and concentration control via the heaviest weight. A main structural result is a tensor-product principle: $S_a(G) = \sum_v d_G(v)^a$ is multiplicative under the tensor (Kronecker) product, implying additivity of E_a and the Nordhaus–Gaddum-type bound $E_a(G) + E_a(\overline{G}) \leqslant 2\log n$ whenever both entropies are defined. Using majorization, we also prove a monotone exponent hierarchy $E_{IP}(G) \geqslant E_{MP}(G) \geqslant E_{YP}(G)$, with equality throughout precisely for regular graphs. To demonstrate chemical relevance, we analyze $N = 3219$ antibacterial compounds curated from the ChEMBL database and show that both classical degree-entropies and Prodeg entropies strongly track established molecular information/complexity measures (BertzCT and $\log_{10}$(Ipc)), while AvgIpc exhibits weaker and nonlinear associations. Finally, we benchmark entropy-only QSPR models for nine physicochemical endpoints using 5-fold cross-validation. Tree ensembles deliver the strongest performance, with particularly high accuracy for size-related properties such as MolMR ($R^2 = 0.970$) and Molecular Weight ($R^2 = 0.945$), whereas MolLogP remains challenging ($R^2 = 0.268$). Across endpoints, the Prodeg block is competitive despite using fewer descriptors, and combining classical and Prodeg entropies yields consistent (typically modest) RMSE gains, supporting Prodeg entropies as compact and interpretable descriptors for chemical graph analysis and QSPR modeling.

ARTICLE IN PRESS

Keywords: Prodeg indices; degree-based graph entropy; Shannon entropy; chemical graph theory; QSPR/QSAR; tensor product of graphs; Nordhaus–Gaddum inequalities; molecular descriptors.

1 Related works

Graph entropy originates from information theory, where Shannon [1] introduced the entropy functional as a quantitative measure of uncertainty for probability distributions. Early links between information, topology, and biological structure were discussed by Rashevsky [2]. A graph-theoretic viewpoint on entropy was developed by Körner in the context of coding and the entropy of graphs [3]. A broad survey of graph entropy and its role in combinatorial optimization was presented by Simonyi [4]. A historical account of major graph entropy measures and their evolution was compiled by Dehmer and Mowshowitz [5], and a comprehensive monograph treatment can be found in the edited volume of Dehmer *et al.* [6]. Further perspectives and developments were summarized by Li and Wei [7].

Degree-based entropies and extremal properties

A common construction in network science and chemical graph theory assigns vertex probabilities via degree-dependent weights and applies Shannon entropy. Cao, Dehmer, and Shi [8] studied extremal properties of degree-based graph entropies and clarified how degree heterogeneity influences entropy values. Ilić [9] investigated extremal values for general degree-based entropy families and provided bounds that depend on the underlying degree-weight model. Ghalavand, Eliasi, and Ashrafi [10] introduced and analyzed a first degree-based entropy variant and derived basic properties and bounds. Dong *et al.* [11] investigated maximum values of degree-based entropies for bipartite graphs. Cambie, Dong, and Mazzamurro [12] refined the bipartite extremal theory and obtained sharp extremal values for degree-based entropies in that setting. Cambie and Mazzamurro [13] resolved Yan's conjecture on the entropy of graphs and strengthened the understanding of minimum-entropy phenomena under size constraints.

Our manuscript belongs to this line by focusing on degree-power weights with exponents $a \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\}$. Unlike studies that treat a single degree-based entropy in isolation, we emphasize a unified framework in which these three entropies share common structural features-regular extremals, product additivity, and complementary bounds-and can therefore be compared systematically.

Degree power sums and Zagreb-type viewpoints

Degree power sums play a central role in extremal graph theory and are often studied under structural constraints. Bollobás and Nikiforov [14] studied degree powers in graphs with forbidden subgraphs and derived extremal bounds in that context. Bollobás and Nikiforov [15] further connected degree power sums to classical extremal results via the Erdős-Stone theorem. In chemical graph theory, classical degree-based indices such as the first Zagreb index were introduced by Gutman and Trinajstić [16], motivating degree-driven descriptors

ARTICLE IN PRESS

and their generalizations. Further developments of Zagreb-type descriptors and their multiplicative variants for constrained graph classes can be found, for example, in the work of Liu *et al.* on Zagreb and multiplicative Zagreb indices of Eulerian graphs [57]. In our setting, the Prodeg-type indices correspond exactly to degree power sums at $\alpha = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, making them natural sources for entropy constructions.

Majorization and inequality tools

Many extremal entropy arguments rely on the Schur-concavity of Shannon entropy and thus on its monotonic decrease under majorization of the induced probability vector. A standard reference for majorization theory and its applications is the book of Marshall and Olkin [17], while inequality tools frequently used in such comparisons trace back to Karamata's inequality [18]. These principles are particularly relevant when comparing degree-based entropy models across families with different degrees of concentration in the underlying degree-weight distribution, and they are also widely used in extremal studies of degree-based graph entropies (see, e.g., [8, 9]). The extremal part of our analysis follows this established methodology, but it is tailored to the Prodeg-type degree-power weights $a \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\}$, yielding sharp attainment characterizations and explicit closed-form evaluations that facilitate a systematic comparison of how different exponents amplify degree heterogeneity.

Other entropy models on graphs

Graph entropy is not restricted to degree-based constructions. Dehmer [19] introduced information functionals to model information processing in complex networks. Chen, Dehmer, and Shi [20] studied distance-based graph entropies and discussed fundamental properties of such measures. Ghorbani, Dehmer, and Zangi [21] investigated graph operations under distance-based entropy models and derived operation-dependent relations. Cao, Dehmer, and Kang [22] proposed network entropies based on independent sets and matchings. Wan *et al.* [23] analyzed entropy measures induced by counts of independent sets and matchings and established further structural results. Wan *et al.* [24] developed graph entropies based on spanning forests of c-cyclic graphs. Related network-information perspectives also arise in structured and recursive networks. In particular, recent studies have examined the statistical and topological properties of fractal networks [60] and the leader–follower coherence and spectral properties of weighted recursive networks under noise disturbance [61]. Coherence-type viewpoints for polygon networks have also been investigated by Liu *et al.* [56].

Chemical graph theory and entropy from topological indices

Topological indices provide a classical bridge between graph structure and chemical properties. In particular, recent studies combine degree-based descriptors with QSPR modeling in diverse molecular settings, including M-polynomial-based constructions for drug-like molecules [58] and descriptor-based prediction of physical properties for carbohydrate derivatives such as guar gum [59].

ARTICLE IN PRESS

Recent studies in chemical graph theory have also developed and applied degree-based topological indices and their coindices-including investigations of disjunction and symmetric-difference operations, complement-graph descriptors, and nanostructure/molecular graph applications-providing further motivation for entropy constructions from degree-powered weights [25, 26, 27, 28]. Wiener [29] introduced the Wiener index to predict boiling points of paraffins. Randić [30] proposed the branching index to quantify molecular branching. Bonchev and Rouvray [31] discussed broader complexity concepts and structural descriptors in chemistry and biology. Manzoor *et al.* [32] studied entropy measures of molecular graphs using topological indices. Manzoor *et al.* [33] computed entropy measures for polycyclic hydroxy-chloroquine structures motivated by COVID-19 applications. Wang *et al.* [34] analyzed entropy measures for silicon carbide networks using degree-based descriptors. Huang *et al.* [35] investigated topological indices and physical analysis via curve fitting for polymer networks, including entropy-related interpretations. Kumar and Das [36] developed Nirmala-index-based entropy measures for silicon carbide networks. Ishfaq *et al.* [37] studied topological indices and entropies for diamond structures.

Positioning of the present work. Motivated by degree-based entropy theory and its applications, we study three Prodeg-type degree-power entropies associated with exponents $\frac{1}{2}, \frac{3}{2}$, and $\frac{5}{2}$. We provide explicit formulas for standard families, establish sharp extremal behavior (with regular graphs as maximizers), and prove tensor-product additivity leading to simple complementary bounds. This product-based viewpoint is related in spirit to graph-product constructions studied in the literature, including the generalized hierarchical product [44] and Hosoya-polynomial investigations of hierarchical products [45]. It also suggests as a natural future direction the study of analogous entropy identities under other graph operations, especially the Cartesian product and related product-type constructions.

We establish a monotone hierarchy $E_{\mathrm{IP}} \geqslant E_{\mathrm{MP}} \geqslant E_{\mathrm{YP}}$, with equality if and only if regular, showing how increasing convexity of degree-weighting forces stronger concentration and lower entropy.

ARTICLE IN PRESS

2 Preliminaries and degree-based indices

2.1 Graph notation

Throughout this work, all graphs are finite, simple (no loops or multiple edges), undirected, and connected. Let $G = (V(G), E(G))$ be a graph of order $n = |V(G)|$ and size $m = |E(G)|$. For $v \in V(G)$, let $d_G(v)$ (or $\deg_G(v)$) denote the degree of v. The minimum and maximum degrees are denoted by $\delta(G)$ and $\Delta(G)$, respectively.

The complement $\overline{G}$ is the graph on $V(G)$ such that $uv \in E(\overline{G})$ if and only if $uv \notin E(G)$. The complete graph K_n has

$$|E(K_n)| = \binom{n}{2} = \frac{n(n-1)}{2},$$

hence

$$|E(G)| + |E(\overline{G})| = \binom{n}{2}, \qquad |E(\overline{G})| = \binom{n}{2} - m.$$

ARTICLE IN PRESS

Moreover, the handshaking lemma gives $\sum_{v\in V(G)} d_G(v) = 2m$, and degrees satisfy

$$d_{\overline{G}}(v) = (n-1) - d_G(v) \qquad \text{for all } v \in V(G).$$

Remark 1. *For any invariant I, the value $I(\overline{G})$ means the index computed on the complement graph. In contrast, $\overline{I}(G)$ denotes a* coindex*, i.e., a sum over* nonedges *of G. We reserve the overline notation $\overline{I}(G)$ exclusively for coindices.*

2.2 Classical degree-based indices

First Zagreb index [16].

$$M_1(G) = \sum_{uv\in E(G)} \big(d_G(u) + d_G(v)\big) = \sum_{v\in V(G)} d_G(v)^2.$$

Forgotten (F) index [38].

$$F(G) = \sum_{uv\in E(G)} \big(d_G(u)^2 + d_G(v)^2\big) = \sum_{v\in V(G)} d_G(v)^3.$$

Yemen (Y) index [39].

$$Y(G) = \sum_{uv\in E(G)} \big(d_G(u)^3 + d_G(v)^3\big) = \sum_{v\in V(G)} d_G(v)^4.$$

Nirmala index [40].

$$N(G) = \sum_{uv\in E(G)} \sqrt{d_G(u) + d_G(v)}.$$

Sombor index [41].

$$SO(G) = \sum_{uv\in E(G)} \sqrt{d_G(u)^2 + d_G(v)^2}.$$

Dharwad index [42].

$$D(G) = \sum_{uv\in E(G)} \sqrt{d_G(u)^3 + d_G(v)^3}.$$

2.3 General Zagreb index and Prodeg-type indices

Definition 2. *For a real parameter α, the* general Zagreb index *[43] is*

$$M_\alpha(G) = \sum_{v\in V(G)} d_G(v)^\alpha.$$

ARTICLE IN PRESS

Remark 3. *Our notation $M_\alpha(G) = \sum_{v\in V(G)} d_G(v)^\alpha$ follows the general degree-power convention. In particular, the classical first Zagreb index satisfies*

$$M_1(G) = \sum_{v\in V(G)} d_G(v)^2 = M_\alpha(G)\big|_{\alpha=2}.$$

Moreover,

$$M_\alpha(G)\big|_{\alpha=1} = \sum_{v\in V(G)} d_G(v) = 2m, \qquad M_\alpha(G)\big|_{\alpha=-1} = \sum_{v\in V(G)} d_G(v)^{-1}.$$

2.4 Inverse Prodeg index and coindex

Definition 4. *Let G be a connected graph of order n. The* Inverse Prodeg index *is*

$$IP(G) = \sum_{uv\in E(G)} \big(d_G(u)^{-1/2} + d_G(v)^{-1/2}\big) = \sum_{v\in V(G)} d_G(v)^{1/2}. \tag{1}$$

The Inverse Prodeg coindex *is*

$$\overline{IP}(G) = \sum_{\{u,v\}\subseteq V(G)\,:\,uv\notin E(G)} \big(d_G(u)^{-1/2} + d_G(v)^{-1/2}\big). \tag{2}$$

Lemma 5. *For any graph G of order n,*

$$\overline{IP}(G) = (n-1)\sum_{v\in V(G)} d_G(v)^{-1/2} - IP(G).$$

Equivalently,

$$\overline{IP}(G) = (n-1)\sum_{v\in V(G)} d_G(v)^{-1/2} - \sum_{v\in V(G)} d_G(v)^{1/2}.$$

Remark 6. *Let $\Delta = \Delta(G)$. Then*

$$IP(G) \geqslant \frac{2m}{\sqrt{\Delta}}, \qquad IP(G) \leqslant \sqrt{2mn},$$

and equality in the upper bound holds if and only if G is regular.

2.5 Misbalance Prodeg index and coindex

Definition 7. *The* Misbalance Prodeg index *is the $\alpha = \frac{3}{2}$ member of M_α:*

$$MP(G) := M_{3/2}(G) = \sum_{v\in V(G)} d_G(v)^{3/2}.$$

For convenience, we also write

$$Z_{3/2}(G) := MP(G).$$

ARTICLE IN PRESS

Lemma 8. *For every graph G,*

$$MP(G) = \sum_{uv \in E(G)} \left(\sqrt{d_G(u)} + \sqrt{d_G(v)}\right).$$

Definition 9. *The square-root degree sum is defined by*

$$Z_{1/2}(G) = \sum_{v \in V(G)} d_G(v)^{1/2}.$$

Remark 10. *Note that $Z_{1/2}(G) = \sum_v \sqrt{d_G(v)}$ coincides with the vertex form of $IP(G)$ in* (1).

Definition 11. *The* Misbalance Prodeg coindex *is the nonedge-sum*

$$\overline{MP}(G) = \sum_{uv \in E(\overline{G})} \left(\sqrt{d_G(u)} + \sqrt{d_G(v)}\right).$$

Proposition 12. *For every graph G of order n,*

$$\overline{MP}(G) = (n-1)\, Z_{1/2}(G) - MP(G).$$

Equivalently,

$$MP(G) + \overline{MP}(G) = (n-1)\, Z_{1/2}(G).$$

2.6 Yemen Prodeg index and coindex

Definition 13. *The* Yemen Prodeg index *is*

$$YP(G) = \sum_{v \in V(G)} d_G(v)^{5/2}.$$

Equivalently, by double counting over incident edges,

$$YP(G) = \sum_{uv \in E(G)} \left(d_G(u)^{3/2} + d_G(v)^{3/2}\right).$$

The Yemen Prodeg coindex *is*

$$\overline{YP}(G) = \sum_{uv \in E(\overline{G})} \left(d_G(u)^{3/2} + d_G(v)^{3/2}\right).$$

Lemma 14. *For any graph G of order n,*

$$\overline{YP}(G) = (n-1)\, Z_{3/2}(G) - YP(G),$$

where $Z_{3/2}(G) = MP(G)$ by Definition 7.

ARTICLE IN PRESS

ARTICLE IN PRESS

2.7 Degree-based entropies

We associate to each vertex-based index $I(G) = \sum_{v\in V(G)} w_G(v)$ a probability distribution

$$p_v = \frac{w_G(v)}{\sum_{u\in V(G)} w_G(u)} \qquad (v \in V(G)),$$

and define the corresponding *degree-based entropy* by the Shannon formula

$$E_I(G) := -\sum_{v\in V(G)} p_v \log p_v,$$

with the convention $0 \log 0 := 0$. Unless stated otherwise, log denotes the natural logarithm.

Definition 15. *Let G be a connected graph of order $n \geqslant 2$. Set*

$$S_{IP}(G) := \sum_{v\in V(G)} \sqrt{d_G(v)} = IP(G), \qquad p_v^{IP} := \frac{\sqrt{d_G(v)}}{S_{IP}(G)}.$$

The Inverse Prodeg entropy *is*

$$E_{IP}(G) := -\sum_{v\in V(G)} p_v^{IP} \log p_v^{IP} = \log IP(G) - \frac{1}{2\,IP(G)} \sum_{v\in V(G)} \sqrt{d_G(v)}\, \log d_G(v). \qquad (3)$$

Definition 16. *Let G be a connected graph of order $n \geqslant 2$. Set*

$$S_{MP}(G) := \sum_{v\in V(G)} d_G(v)^{3/2} = MP(G), \qquad p_v^{MP} := \frac{d_G(v)^{3/2}}{S_{MP}(G)}.$$

The Misbalance Prodeg entropy *is*

$$E_{MP}(G) := -\sum_{v\in V(G)} p_v^{MP} \log p_v^{MP} = \log MP(G) - \frac{3}{2\,MP(G)} \sum_{v\in V(G)} d_G(v)^{3/2}\, \log d_G(v). \qquad (4)$$

Definition 17. *Let G be a connected graph of order $n \geqslant 2$. Set*

$$S_{YP}(G) := \sum_{v\in V(G)} d_G(v)^{5/2} = YP(G), \qquad p_v^{YP} := \frac{d_G(v)^{5/2}}{S_{YP}(G)}.$$

The Yemen Prodeg entropy *is*

$$E_{YP}(G) := -\sum_{v\in V(G)} p_v^{YP} \log p_v^{YP} = \log YP(G) - \frac{5}{2\,YP(G)} \sum_{v\in V(G)} d_G(v)^{5/2}\, \log d_G(v). \qquad (5)$$

Proposition 18. *Let G be a connected graph of order $n \geqslant 2$. Then*

$$E_{IP}(G) \leqslant \log n, \qquad E_{MP}(G) \leqslant \log n, \qquad E_{YP}(G) \leqslant \log n.$$

Moreover, equality in any of these holds if and only if G is regular (equivalently, the corresponding weights are uniform).

Remark 19. *For connected graphs, all degrees are positive, hence the above entropies are well-defined and strictly positive. In particular, E_{MP} and E_{YP} can be arbitrarily small on highly imbalanced families (e.g., stars as $n \to \infty$), whereas E_{IP} is much less concentrated due to the sublinear weighting $\sqrt{d}$.*

ARTICLE IN PRESS

ARTICLE IN PRESS

2.8 Tensor-product additivity for degree-based entropies

Definition 20. *Given graphs G_1 and G_2, their tensor product $G_1 \otimes G_2$ has vertex set $V(G_1) \times V(G_2)$, where $(u_1, v_1) \sim (u_2, v_2)$ if and only if $u_1u_2 \in E(G_1)$ and $v_1v_2 \in E(G_2)$. Degrees satisfy*

$$d_{G_1 \otimes G_2}(u, v) = d_{G_1}(u)\, d_{G_2}(v). \tag{6}$$

Remark 21. *Graph products and related hierarchical constructions have been studied from several combinatorial viewpoints; see, for example, the generalized hierarchical product of graphs [44] and the Hosoya polynomial of hierarchical product graphs [45]. In the present work, we focus on the tensor (Kronecker) product because the multiplicative degree identity in* (6) *leads directly to exact additivity of the degree-power entropies. Extending this analysis to other graph operations, especially the Cartesian product and related product-type constructions, is a natural direction for future study.*

Lemma 22. *Let $a > 0$ and define $S_a(G) := \sum_{v \in V(G)} d_G(v)^a$. Then for any graphs G_1, G_2,*

$$S_a(G_1 \otimes G_2) = S_a(G_1)\, S_a(G_2).$$

Proof. By (6),

$$S_a(G_1 \otimes G_2) = \sum_{(u,v) \in V(G_1) \times V(G_2)} \big(d_{G_1}(u) d_{G_2}(v)\big)^a = \sum_u \sum_v d_{G_1}(u)^a\, d_{G_2}(v)^a = S_a(G_1)\, S_a(G_2).$$

□

Theorem 23. *Fix $a > 0$ and define degree-based weights*

$$p_v^{(a)}(G) := \frac{d_G(v)^a}{S_a(G)}, \qquad S_a(G) = \sum_{u \in V(G)} d_G(u)^a,$$

assuming $S_a(G) > 0$ (equivalently, G has at least one edge). Let

$$E_a(G) := -\sum_{v \in V(G)} p_v^{(a)}(G)\, \log p_v^{(a)}(G)$$

be the associated Shannon entropy. Then, for any graphs G_1, G_2 with $S_a(G_i) > 0$,

$$E_a(G_1 \otimes G_2) = E_a(G_1) + E_a(G_2).$$

In particular,

$$E_{IP}(G_1 \otimes G_2) = E_{IP}(G_1) + E_{IP}(G_2), \quad E_{MP}(G_1 \otimes G_2) = E_{MP}(G_1) + E_{MP}(G_2),$$

$$E_{YP}(G_1 \otimes G_2) = E_{YP}(G_1) + E_{YP}(G_2),$$

corresponding to $a = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$, respectively.

ARTICLE IN PRESS

Proof. By (6) and Lemma 22, we have

$$p^{(a)}_{(u,v)}(G_1 \otimes G_2) = \frac{\big(d_{G_1}(u)d_{G_2}(v)\big)^a}{S_a(G_1 \otimes G_2)} = \frac{d_{G_1}(u)^a}{S_a(G_1)} \cdot \frac{d_{G_2}(v)^a}{S_a(G_2)} = p^{(a)}_u(G_1)\, p^{(a)}_v(G_2).$$

Hence the distribution on $V(G_1 \otimes G_2)$ is the product distribution $p^{(a)}(G_1) \otimes p^{(a)}(G_2)$. Shannon entropy is additive on product distributions, so

$$E_a(G_1 \otimes G_2) = E_a(G_1) + E_a(G_2).$$

Substituting $a = \frac{1}{2}, \frac{3}{2}, \frac{5}{2}$ yields the three stated cases. □

The exact additivity above relies essentially on the multiplicative degree structure of the tensor product; analogous entropy identities for other products, such as the Cartesian product, remain open and will be investigated in future work.

Corollary 24. *Let $a > 0$ and let G satisfy $S_a(G) > 0$. For any integer $k \geqslant 1$,*

$$E_a\big(G^{\otimes k}\big) = k\, E_a(G).$$

In particular,

$$E_{IP}\big(G^{\otimes k}\big) = k\, E_{IP}(G), \quad E_{MP}\big(G^{\otimes k}\big) = k\, E_{MP}(G), \quad E_{YP}\big(G^{\otimes k}\big) = k\, E_{YP}(G).$$

Proof. Apply Theorem 23 repeatedly (induction on k). □

Remark 25. *If G is edgeless, then $d_G(v) = 0$ for all v, hence $S_a(G) = 0$ and the weights are undefined. For connected graphs with $|V(G)| \geqslant 2$, we automatically have $S_a(G) > 0$.*

Corollary 26. *For any graph G of order n for which both entropies are defined,*

$$E_a(G) + E_a(\overline{G}) \leqslant 2 \log n.$$

In particular,

$$E_{IP}(G) + E_{IP}(\overline{G}) \leqslant 2 \log n, \quad E_{MP}(G) + E_{MP}(\overline{G}) \leqslant 2 \log n, \quad E_{YP}(G) + E_{YP}(\overline{G}) \leqslant 2 \log n.$$

Proof. Each entropy is at most $\log n$ (maximum entropy on n atoms). Add the two bounds. □

3 Inverse Prodeg entropy

In this section we study the Inverse Prodeg entropy $E_{IP}(G)$ introduced in Definition 15. Throughout, G is a simple, finite, connected graph of order $n \geqslant 2$.

ARTICLE IN PRESS

3.1 Representative exact formulas

Proposition 27.

1. *If G is d-regular (in particular K_n and C_n), then $p_v^{IP} \equiv 1/n$ and*

$$E_{IP}(G) = \log n.$$

2. *For the star S_n ($n \geqslant 2$), with hub weight p_c and leaf weight p_ℓ,*

$$E_{IP}(S_n) = -\big(p_c \log p_c + (n-1) p_\ell \log p_\ell\big), \qquad p_c = \frac{\sqrt{n-1}}{\sqrt{n-1}+n-1}, \quad p_\ell = \frac{1}{\sqrt{n-1}+n-1}. \tag{7}$$

3. *For the path P_n ($n \geqslant 2$), with two endpoints of degree 1 and $n-2$ internal vertices of degree 2,*

$$E_{IP}(P_n) = -\big(2p_1 \log p_1 + (n-2) p_2 \log p_2\big), \qquad p_1 = \frac{1}{2+(n-2)\sqrt{2}}, \quad p_2 = \frac{\sqrt{2}}{2+(n-2)\sqrt{2}}. \tag{8}$$

4. *For the complete bipartite graph $K_{m,n}$ (order $m+n$),*

$$E_{IP}(K_{m,n}) = -\big(mp_a \log p_a + np_b \log p_b\big), \qquad p_a = \frac{\sqrt{n}}{m\sqrt{n}+n\sqrt{m}}, \quad p_b = \frac{\sqrt{m}}{m\sqrt{n}+n\sqrt{m}}. \tag{9}$$

Table 1: Inverse Prodeg entropy: exact values and bounds for common classes

Graph class	**Exact $E_{IP}(G)$**	**Bounds**
Complete K_n	$\log n$	-
Cycle C_n	$\log n$	-
Star S_n	(7)	$\log 2 < E_{IP}(S_n) < \log n$ $(n \geqslant 3)$
Path P_n	(8)	$\log 2 < E_{IP}(P_n) < \log n$ $(n \geqslant 3)$
Tree T_n	-	$\log 2 < E_{IP}(T_n) < \log n$ $(n \geqslant 3)$
Complete bipartite $K_{m,n}$	(9)	$\log 2 \leqslant E_{IP}(K_{m,n}) \leqslant \log(m+n)$
d-regular	$\log n$	-
Single edge K_2	$\log 2$	-

From Table 1, the maximum value $\log n$ is attained precisely by regular graphs. The minimum among connected graphs is $\log 2$, attained by K_2.

3.2 Bounds and extremal graphs

As in the general theory of degree-based Shannon entropies, our extremal analysis is guided by the Schur-concavity of Shannon entropy and majorization arguments for the induced

ARTICLE IN PRESS

probability vectors; see, for example, [17, 8, 9]. In the present setting, these ideas yield explicit closed-form evaluations for representative graph families, a sharp characterization of the equality case in the upper bound $\log n$, and a systematic comparison across the three Prodeg-type entropy models ($a \in \{\frac{1}{2}, \frac{3}{2}, \frac{5}{2}\}$).

Theorem 28. *Let G be a connected graph on $n \geqslant 2$ vertices. Then*

$$\log 2 \;\leqslant\; E_{IP}(G) \;\leqslant\; \log n.$$

Moreover, $E_{IP}(G) = \log n$ if and only if G is regular. The lower bound $\log 2$ is attained by K_2.

Proof. Let $p_v^{IP} = \frac{\sqrt{d_G(v)}}{\sum_{u \in V(G)} \sqrt{d_G(u)}}$.

Upper bound. For any distribution on n points, Shannon entropy is at most $\log n$, with equality iff the distribution is uniform. Here uniformity means all $\sqrt{d_G(v)}$ are equal, hence $d_G(v)$ is constant and G is regular.

Lower bound. Let $\Delta = \Delta(G)$ and choose v^* with $d_G(v^*) = \Delta$. Since G is connected, $d_G(v) \geqslant 1$ for all v, so

$$\sum_{u \in V(G)} \sqrt{d_G(u)} \;\geqslant\; \sqrt{\Delta} + (n-1).$$

Hence

$$p_{\max} = \max_v p_v^{IP} \leqslant \frac{\sqrt{\Delta}}{\sqrt{\Delta} + (n-1)} \leqslant \frac{\sqrt{n-1}}{\sqrt{n-1} + (n-1)} = \frac{1}{1 + \sqrt{n-1}} \leqslant \frac{1}{2}.$$

For any distribution, $H(p) \geqslant -\log p_{\max}$, hence $E_{IP}(G) \geqslant \log 2$. Equality holds for K_2. □

Remark 29. *Tensor additivity of E_{IP} is the case $a = \frac{1}{2}$ of Theorem 23.*

Corollary 30. *This is the case $a = \frac{1}{2}$ of Corollary 26:*

$$E_{IP}(G) + E_{IP}(\overline{G}) \leqslant 2 \log n,$$

whenever both entropies are defined.

A useful increment under edge addition

Remark 31. *If G' is obtained from G by adding a new edge xy, then*

$$IP(G') - IP(G) = \big(\sqrt{d_G(x) + 1} - \sqrt{d_G(x)}\big) + \big(\sqrt{d_G(y) + 1} - \sqrt{d_G(y)}\big),$$

which follows directly from the vertex form $IP(G) = \sum_{v \in V(G)} \sqrt{d_G(v)}$.

4 Misbalance Prodeg entropy

In this section we study the Misbalance Prodeg entropy $E_{MP}(G)$ defined in Definition 16. Throughout, graphs are connected unless stated otherwise, so $d_G(v) \geqslant 1$ and $MP(G) > 0$.

ARTICLE IN PRESS

4.1 Exact formulas for common graph classes

Proposition 32. *Let G be a connected graph of order n.*

1. *If G is d-regular, then $p_v^{MP} \equiv 1/n$ and*

$$E_{MP}(G) = \log n.$$

2. *For the star S_n ($n \geqslant 2$),*

$$E_{MP}(S_n) = -\big(p_c \log p_c + (n-1) p_\ell \log p_\ell\big), \tag{10}$$

where

$$p_c = \frac{(n-1)^{3/2}}{(n-1)^{3/2} + (n-1)}, \qquad p_\ell = \frac{1}{(n-1)^{3/2} + (n-1)}.$$

3. *For the path P_n ($n \geqslant 2$),*

$$E_{MP}(P_n) = -\big(2p_1 \log p_1 + (n-2) p_2 \log p_2\big), \tag{11}$$

with

$$p_1 = \frac{1}{2 + 2\sqrt{2}(n-2)}, \qquad p_2 = \frac{2\sqrt{2}}{2 + 2\sqrt{2}(n-2)}.$$

4. *For the complete bipartite graph $K_{r,s}$ ($r, s \geqslant 1$),*

$$E_{MP}(K_{r,s}) = -\big(r\, p_a \log p_a + s\, p_b \log p_b\big), \tag{12}$$

where

$$p_a = \frac{s^{3/2}}{r\, s^{3/2} + s\, r^{3/2}}, \qquad p_b = \frac{r^{3/2}}{r\, s^{3/2} + s\, r^{3/2}}.$$

ARTICLE IN PRESS

Table 2: Misbalance Prodeg entropy: exact values and standard comments

Graph class	**Exact $E_{MP}(G)$**	**Remarks**
Complete K_n	$\log n$	regular
Cycle C_n	$\log n$	regular
d-regular graph	$\log n$	regular
Single edge K_2	$\log 2$	$n = 2$
Star S_n	Eq. (10)	$E_{MP}(S_n) \to 0$ as $n \to \infty$
Path P_n	Eq. (11)	$E_{MP}(P_n) = \log n + o(1)$
Complete bipartite $K_{r,s}$	Eq. (12)	$\log(r+s)$ iff $r = s$ (regular)
Tree T_n	-	$0 < E_{MP}(T_n) \leqslant \log n$

Remark 33. *Regular graphs maximize the entropy: $E_{MP}(G) = \log n$ iff G is regular (Proposition 18). For paths, the normalized $d^{3/2}$-weights become asymptotically uniform, hence $E_{MP}(P_n) \to \log n$ as $n \to \infty$. For stars, the weights concentrate at the hub and $E_{MP}(S_n) \to 0$ as $n \to \infty$.*

ARTICLE IN PRESS

4.2 Bounds in terms of MP, δ, and Δ

Theorem 34. *Let G be a connected graph with minimum and maximum degrees $\delta = \delta(G) \geqslant 1$ and $\Delta = \Delta(G)$. Then*

$$\log MP(G) - \frac{3}{2}\log \Delta \;\leqslant\; E_{MP}(G) \;\leqslant\; \log MP(G) - \frac{3}{2}\log \delta. \tag{13}$$

Equality in either bound holds if and only if G is regular.

Proof. Let $p_v^{MP} = d_G(v)^{3/2}/MP(G)$. From (4),

$$E_{MP}(G) = \log MP(G) - \frac{3}{2}\sum_{v\in V(G)} p_v^{MP}\,\log d_G(v).$$

Since $\delta \leqslant d_G(v) \leqslant \Delta$ for all v, we have

$$\log \delta \;\leqslant\; \sum_v p_v^{MP}\log d_G(v) \;\leqslant\; \log \Delta,$$

which yields (13). Equality forces $\log d_G(v)$ to be constant on $V(G)$, hence G is regular. □

Corollary 35. *Let G be a connected graph of order n, size m, and degree extremes δ, Δ. Then*

$$E_{MP}(G) \geqslant \log\left(\frac{(2m)^{3/2}}{\sqrt{n}}\right) - \frac{3}{2}\log\Delta = \frac{3}{2}\log\left(\frac{2m}{\Delta}\right) - \frac{1}{2}\log n, \tag{14}$$

$$E_{MP}(G) \leqslant \log n. \tag{15}$$

Proof. The lower bound follows from the Jensen/Cauchy-Schwarz estimate $MP(G) \geqslant (2m)^{3/2}/\sqrt{n}$ together with the left inequality in (13). The upper bound is the universal entropy bound on n atoms; see Proposition 18. □

4.3 Majorization viewpoint and extremal trees

Theorem 36. *Let $\mathbf{p}, \mathbf{q}$ be probability vectors on n atoms. If $\mathbf{p} \succ \mathbf{q}$, then $H(\mathbf{p}) \leqslant H(\mathbf{q})$. In particular, if the MP-weight vector $\mathbf{p}^{MP}(G)$ majorizes $\mathbf{p}^{MP}(H)$, then $E_{MP}(G) \leqslant E_{MP}(H)$.*

Conjecture 1. *Among all trees T on $n \geqslant 2$ vertices,*

$$E_{MP}(S_n) \;\leqslant\; E_{MP}(T) \;\leqslant\; E_{MP}(P_n).$$

Remark 37. *This conjecture reflects the natural expectation that the $d^{3/2}$-weights are most concentrated for the star and most evenly distributed for the path. A rigorous proof would require a systematic comparison of the normalized MP-weight vectors over the class of all trees.*

ARTICLE IN PRESS

ARTICLE IN PRESS

Asymptotics for Erdős-Rényi random graphs

Proposition 38. *Let $G \sim G(n,p)$ with fixed $p \in (0,1)$, and let $D \sim \mathrm{Bin}(n-1,p)$ be the degree of a fixed vertex. Then, as $n \to \infty$,*

$$\mathbb{E}[E_{MP}(G)] = \log n - \frac{9}{8}\frac{\mathrm{Var}(D)}{(\mathbb{E}D)^2} + o\left(\frac{1}{n}\right) = \log n - \frac{9}{8}\frac{1-p}{p}\cdot\frac{1}{n} + o\left(\frac{1}{n}\right).$$

Remark 39. *The above expansion is heuristic. It is obtained by perturbing the MP-weights around the uniform distribution and combining a second-order Taylor expansion of Shannon entropy with a first-order expansion of $x^{3/2}$ around $\mathbb{E}D$. A fully rigorous derivation would require concentration estimates together with uniform control of the remainder terms.*

Heuristic proof. Expand the MP-weights as a small perturbation of the uniform distribution and apply a second-order Taylor expansion of Shannon entropy around uniformity, together with a first-order expansion of $x^{3/2}$ around $\mathbb{E}D$; this yields the stated correction term of order $1/n$. □

Behavior under graph operations (regular factors)

Remark 40. *Whenever a graph operation produces a regular graph on N vertices (e.g., products of regular factors that preserve regularity), the MP-weights are uniform and $E_{MP} = \log N$. In particular, for the hypercube $Q_d = K_2^{\square d}$ one has $E_{MP}(Q_d) = \log(2^d) = d\log 2$.*

Remark 41. *If G' is obtained from G by adding a new edge xy, then only $d_G(x), d_G(y)$ increase by 1, and*

$$MP(G') - MP(G) = \big((d_G(x)+1)^{3/2} - d_G(x)^{3/2}\big) + \big((d_G(y)+1)^{3/2} - d_G(y)^{3/2}\big).$$

ARTICLE IN PRESS

5 Yemen Prodeg entropy

In this section we study the Yemen Prodeg entropy $E_{YP}(G)$ introduced in Definition 17. Throughout, graphs are connected, hence $d_G(v) \geqslant 1$ and $YP(G) > 0$.

5.1 Basic bounds and extremals

Theorem 42. *Let G be a connected graph on $n \geqslant 2$ vertices. Then*

$$0 \;<\; E_{YP}(G) \;\leqslant\; \log n.$$

Moreover, $E_{YP}(G) = \log n$ if and only if G is regular (equivalently, $p_v^{YP} \equiv 1/n$). For $n = 2$ (i.e., $G \cong K_2$), one has $E_{YP}(G) = \log 2$. For $n \geqslant 3$, the entropy can be arbitrarily small (e.g., $E_{YP}(S_n) \to 0$ as $n \to \infty$).

Proof. The upper bound $E_{YP}(G) \leqslant \log n$ and the characterization of equality are immediate from Proposition 18. Since G is connected, all degrees are positive, so all YP-weights $p_v^{YP} = d_G(v)^{5/2}/YP(G)$ are strictly positive and therefore $E_{YP}(G) > 0$. Finally, the star family satisfies $E_{YP}(S_n) \to 0$ by the explicit formula in Proposition 44 below. □

ARTICLE IN PRESS

Lemma 43. *Let G be connected of order n and let $\Delta = \Delta(G)$. Then*

$$E_{YP}(G) \;\geqslant\; -\log p_{\max} \;\geqslant\; \log\Big(1 + \frac{n-1}{\Delta^{5/2}}\Big),$$

where $p_{\max} = \max_{v \in V(G)} p_v^{YP}$.

Proof. Since G is connected, every vertex has degree at least 1, hence

$$YP(G) = \sum_{v \in V(G)} d_G(v)^{5/2} \;\geqslant\; \Delta^{5/2} + (n-1) \cdot 1 = \Delta^{5/2} + (n-1).$$

Therefore

$$p_{\max} = \frac{\Delta^{5/2}}{YP(G)} \;\leqslant\; \frac{\Delta^{5/2}}{\Delta^{5/2} + (n-1)}.$$

Also Shannon entropy dominates the min-entropy: $H(p) \geqslant -\log p_{\max}$. Combining yields the claim. □

5.2 Exact formulas for representative families

Proposition 44. *Let G be a connected graph of order n.*

1. *If G is d-regular (e.g. K_n, C_n, Q_d), then $p_v^{YP} \equiv 1/n$ and*

$$E_{YP}(G) = \log n.$$

2. *For the star S_n ($n \geqslant 2$), with hub degree $n-1$ and leaf degree 1,*

$$E_{YP}(S_n) = -\Big(p_c \log p_c + (n-1) p_\ell \log p_\ell\Big),$$
$$p_c = \frac{(n-1)^{5/2}}{(n-1)^{5/2} + (n-1)}, \qquad p_\ell = \frac{1}{(n-1)^{5/2} + (n-1)}. \tag{16}$$

   In particular, $E_{YP}(S_n) \downarrow 0$ as $n \to \infty$.

3. *For the path P_n ($n \geqslant 2$), with two ends of degree 1 and $n-2$ internal vertices of degree 2,*

$$E_{YP}(P_n) = -\Big(2p_1 \log p_1 + (n-2) p_2 \log p_2\Big),$$
$$p_1 = \frac{1}{2 + 4\sqrt{2}\,(n-2)}, \qquad p_2 = \frac{4\sqrt{2}}{2 + 4\sqrt{2}\,(n-2)}. \tag{17}$$

   Moreover, $E_{YP}(P_n) = \log n + o(1)$ as $n \to \infty$.

4. *For the complete bipartite graph $K_{r,s}$ ($r, s \geqslant 1$),*

$$E_{YP}(K_{r,s}) = -\Big(r\, p_a \log p_a + s\, p_b \log p_b\Big),$$
$$p_a = \frac{s^{5/2}}{r\, s^{5/2} + s\, r^{5/2}}, \qquad p_b = \frac{r^{5/2}}{r\, s^{5/2} + s\, r^{5/2}}. \tag{18}$$

   In particular, $E_{YP}(K_{r,s}) = \log(r+s)$ if and only if $r = s$ (the regular case).

ARTICLE IN PRESS

ARTICLE IN PRESS

Remark 45. *For paths, $E_{YP}(P_n)$ increases with n and satisfies $E_{YP}(P_n) = \log n + o(1)$, whereas for stars, $E_{YP}(S_n)$ decreases with n and tends to 0. Table 3 summarizes representative exact values and standard bounds.*

Table 3: YP-entropy: exact values and standard bounds for common classes

Graph class	**Exact $E_{YP}(G)$**	**Bounds / remarks**
Complete K_n	$\log n$	regular
Cycle C_n	$\log n$	regular
d-regular graph	$\log n$	regular
Single edge K_2	$\log 2$	$n = 2$
Star S_n	Eq. (16)	$0 < E_{YP}(S_n) < \log n$, $E_{YP}(S_n) \to 0$
Path P_n	Eq. (17)	$0 < E_{YP}(P_n) < \log n$, $E_{YP}(P_n) = \log n + o(1)$
Complete bipartite $K_{r,s}$	Eq. (18)	$\leqslant \log(r+s)$, equality iff $r = s$
Tree T_n	-	$0 < E_{YP}(T_n) \leqslant \log n$

5.3 A concentration-type upper bound

Remark 46. *Let $S = YP(G) = \sum_{v \in V(G)} d_G(v)^{5/2}$ and*

$$p_{\max} = \max_{v \in V(G)} \frac{d_G(v)^{5/2}}{S} = \frac{\Delta(G)^{5/2}}{YP(G)}.$$

Among all probability distributions on n points with fixed $p_{\max}$, Shannon entropy is maximized when the remaining mass is distributed uniformly over the other $n-1$ points. Hence

$$E_{YP}(G) \;\leqslant\; h(p_{\max}) + (1 - p_{\max}) \log(n-1), \qquad h(t) := -t \log t - (1-t) \log(1-t). \quad (19)$$

Equivalently,

$$E_{YP}(G) \;\leqslant\; h\Big(\frac{\Delta(G)^{5/2}}{YP(G)}\Big) + \Big(1 - \frac{\Delta(G)^{5/2}}{YP(G)}\Big) \log(n-1). \quad (20)$$

Together with Lemma 43, this yields the two-sided estimate

$$\log\Big(1 + \frac{n-1}{\Delta(G)^{5/2}}\Big) \;\leqslant\; E_{YP}(G) \;\leqslant\; h\Big(\frac{\Delta(G)^{5/2}}{YP(G)}\Big) + \Big(1 - \frac{\Delta(G)^{5/2}}{YP(G)}\Big) \log(n-1).$$

Remark 47. *The tensor additivity of E_{YP} is a special case of Theorem 23 (with $a = \frac{5}{2}$), and the complementary bound $E_{YP}(G) + E_{YP}(\overline{G}) \leqslant 2 \log n$ follows from Corollary 26.*

Remark 48. *If G' is obtained from G by adding a new edge xy, then only the degrees of x and y change, and*

$$YP(G') - YP(G) = \big((d_G(x)+1)^{5/2} - d_G(x)^{5/2}\big) + \big((d_G(y)+1)^{5/2} - d_G(y)^{5/2}\big).$$

ARTICLE IN PRESS

6 Hierarchy and comparison principles

Throughout this section, G denotes a connected graph of order $n \geqslant 2$. All logarithms are natural unless stated otherwise. For a real parameter $a > 0$, define the degree power sum

$$S_a(G) := \sum_{v \in V(G)} d_G(v)^a, \tag{21}$$

the associated probability weights

$$p_v^{(a)}(G) := \frac{d_G(v)^a}{S_a(G)} \qquad (v \in V(G)), \tag{22}$$

and the degree-power Shannon entropy

$$E_a(G) := -\sum_{v \in V(G)} p_v^{(a)}(G) \log p_v^{(a)}(G). \tag{23}$$

In particular,

$$E_{IP}(G) = E_{1/2}(G), \qquad E_{MP}(G) = E_{3/2}(G), \qquad E_{YP}(G) = E_{5/2}(G).$$

Equivalently, expanding $-\sum p \log p$ and using $\sum_v p_v^{(a)}(G) = 1$ gives the identity

$$E_a(G) = \log S_a(G) - \frac{a}{S_a(G)} \sum_{v \in V(G)} d_G(v)^a \log d_G(v). \tag{24}$$

6.1 Exponent hierarchy via majorization

Theorem 49. *Let $0 < a < b$ and let $d_1 \geqslant d_2 \geqslant \cdots \geqslant d_n \geqslant 1$ be the degree sequence of G in nonincreasing order. Define*

$$p_i^{(r)} := \frac{d_i^r}{\sum_{j=1}^n d_j^r} \qquad (r > 0,\ i = 1, \ldots, n).$$

Then the probability vector $\mathbf{p}^{(b)} = (p_1^{(b)}, \ldots, p_n^{(b)})$ majorizes $\mathbf{p}^{(a)} = (p_1^{(a)}, \ldots, p_n^{(a)})$, i.e.,

$$\mathbf{p}^{(b)} \ \succ \ \mathbf{p}^{(a)}.$$

Consequently, since Shannon entropy is Schur-concave (see, e.g., [17]),

$$E_a(G) \ \geqslant \ E_b(G), \tag{25}$$

with equality if and only if G is regular.

Intuition. The exponent r controls how strongly high degrees are emphasized in the normalized weights $p_i^{(r)} \propto d_i^r$. When r increases, large degrees are magnified relative to small ones, so a larger fraction of the total weight shifts toward the highest-degree vertices. In other words, the distribution $\mathbf{p}^{(r)}$ becomes more *concentrated* as r grows, which is exactly the content of the majorization relation $\mathbf{p}^{(b)} \succ \mathbf{p}^{(a)}$ for $0 < a < b$. Since Shannon entropy is Schur-concave, greater concentration yields smaller entropy, giving the hierarchy $E_a(G) \geqslant E_b(G)$. Equality occurs precisely when the degree sequence is constant (i.e., G is regular), in which case every $\mathbf{p}^{(r)}$ is uniform.

ARTICLE IN PRESS

Proof. For $r > 0$ and $k \in \{1, \dots, n\}$, set

$$A_k(r) := \sum_{i=1}^{k} d_i^r, \qquad A_n(r) := \sum_{i=1}^{n} d_i^r = S_r(G),$$

and define the partial sums

$$s_k(r) := \sum_{i=1}^{k} p_i^{(r)} = \frac{A_k(r)}{A_n(r)}.$$

It suffices to show that $s_k(r)$ is nondecreasing in r for each $k = 1, \dots, n-1$.

Differentiate $\log s_k(r) = \log A_k(r) - \log A_n(r)$:

$$\frac{d}{dr} \log A_k(r) = \sum_{i=1}^{k} \frac{d_i^r}{A_k(r)} \log d_i =: \mu_k(r), \qquad \frac{d}{dr} \log A_n(r) = \sum_{i=1}^{n} \frac{d_i^r}{A_n(r)} \log d_i =: \mu_n(r).$$

Hence

$$\frac{d}{dr} \log s_k(r) = \mu_k(r) - \mu_n(r).$$

To justify the comparison of these two logarithmic means, write

$$B_k(r) := A_n(r) - A_k(r) = \sum_{i=k+1}^{n} d_i^r$$

and, for $k < n$, define the weighted logarithmic mean over the remaining vertices by

$$\nu_k(r) := \sum_{i=k+1}^{n} \frac{d_i^r}{B_k(r)} \log d_i.$$

Then

$$\mu_n(r) = \frac{A_k(r)}{A_n(r)} \mu_k(r) + \frac{B_k(r)}{A_n(r)} \nu_k(r).$$

Because the degree sequence is ordered as $d_1 \geqslant \cdots \geqslant d_k \geqslant d_{k+1} \geqslant \cdots \geqslant d_n$, every logarithm appearing in the first weighted mean is at least every logarithm appearing in the second weighted mean; hence $\mu_k(r) \geqslant \nu_k(r)$. Consequently,

$$\mu_n(r) \leqslant \frac{A_k(r)}{A_n(r)} \mu_k(r) + \frac{B_k(r)}{A_n(r)} \mu_k(r) = \mu_k(r),$$

and therefore $\frac{d}{dr} \log s_k(r) = \mu_k(r) - \mu_n(r) \geqslant 0$. Thus $s_k(r)$ is nondecreasing in r for every $k = 1, \dots, n-1$, which proves $\mathbf{p}^{(b)} \succ \mathbf{p}^{(a)}$. The entropy inequality (25) follows from the Schur-concavity of Shannon entropy. If G is regular, all vectors $\mathbf{p}^{(r)}$ are uniform and equality is immediate. Conversely, if G is not regular, the ordered degree sequence has at least one strict drop; for the corresponding partial sum the above comparison is strict on any interval $a < r < b$, so $\mathbf{p}^{(b)}$ strictly majorizes $\mathbf{p}^{(a)}$. Since Shannon entropy is strictly Schur-concave on the probability simplex, equality in (25) cannot occur. Hence equality holds if and only if $d_1 = \cdots = d_n$, i.e., G is regular. □

ARTICLE IN PRESS

Corollary 50. *For every connected graph G on $n \geqslant 2$ vertices,*

$$E_{IP}(G) = E_{1/2}(G) \;\geqslant\; E_{3/2}(G) = E_{MP}(G) \;\geqslant\; E_{5/2}(G) = E_{YP}(G).$$

Moreover, equality holds throughout if and only if G is regular, in which case

$$E_{IP}(G) = E_{MP}(G) = E_{YP}(G) = \log n \quad (\textit{cf. } [1]).$$

Corollary 51. *Let G be connected. Any lower bound for $E_{YP}(G)$ is automatically a lower bound for $E_{MP}(G)$ and $E_{IP}(G)$:*

$$E_{YP}(G) \geqslant L \implies E_{MP}(G) \geqslant L \quad \textit{and} \quad E_{IP}(G) \geqslant L.$$

Any upper bound for $E_{IP}(G)$ is automatically an upper bound for $E_{MP}(G)$ and $E_{YP}(G)$:

$$E_{IP}(G) \leqslant U \implies E_{MP}(G) \leqslant U \quad \textit{and} \quad E_{YP}(G) \leqslant U.$$

Remark 52. *A second route to (25) follows from (24). Writing $S(a) = S_a(G)$ and $p_v(a) = p_v^{(a)}(G)$, define $\mu(a) := \sum_v p_v(a) \log d_G(v)$. Then $E_a(G) = \log S(a) - a\mu(a)$ and*

$$\frac{d}{da} E_a(G) = -a \operatorname{Var}_{p(a)}\big(\log d_G(v)\big) \leqslant 0,$$

so $E_a(G)$ is nonincreasing in a, with equality only when G is regular.

6.2 Concentration control via the heaviest weight

Lemma 53. *Let G be connected on n vertices and let $a > 0$. Put*

$$p_{\max}^{(a)} := \max_{v \in V(G)} p_v^{(a)}(G) = \frac{\Delta(G)^a}{S_a(G)}.$$

Then

$$-\log p_{\max}^{(a)} \;\leqslant\; E_a(G) \;\leqslant\; h\big(p_{\max}^{(a)}\big) + (1 - p_{\max}^{(a)}) \log(n-1),$$

where $h(t) = -t \log t - (1-t)\log(1-t)$ is the binary entropy function (with the same log base as above).

6.3 Edge-addition increments and hub sensitivity

Remark 54. *Fix $a > 0$ and write $S_a(G) = \sum_v d_G(v)^a$. If G' is obtained from G by adding a new edge xy, then only the degrees of x and y change, hence*

$$S_a(G') - S_a(G) = \Big((d_G(x)+1)^a - d_G(x)^a\Big) + \Big((d_G(y)+1)^a - d_G(y)^a\Big).$$

In particular, for the three Prodeg sums

$$IP(G) = S_{1/2}(G), \qquad MP(G) = S_{3/2}(G), \qquad YP(G) = S_{5/2}(G),$$

we obtain the explicit increments

$$\begin{aligned}
IP(G') - IP(G) &= \big(\sqrt{d_G(x)+1} - \sqrt{d_G(x)}\big) + \big(\sqrt{d_G(y)+1} - \sqrt{d_G(y)}\big), \\
MP(G') - MP(G) &= \big((d_G(x)+1)^{3/2} - d_G(x)^{3/2}\big) + \big((d_G(y)+1)^{3/2} - d_G(y)^{3/2}\big), \\
YP(G') - YP(G) &= \big((d_G(x)+1)^{5/2} - d_G(x)^{5/2}\big) + \big((d_G(y)+1)^{5/2} - d_G(y)^{5/2}\big).
\end{aligned}$$

ARTICLE IN PRESS

Remark 55. *Let $\Delta_a(t) := (t+1)^a - t^a$ for integers $t \geqslant 1$. Then $\Delta_a(t)$ is decreasing in t for $0 < a < 1$ (concave regime) and increasing in t for $a > 1$ (convex regime). Consequently:*

- *for IP ($a = \frac{1}{2}$), adding an edge to already-large degrees yields a smaller increase in IP;*
- *for MP and YP ($a = \frac{3}{2}, \frac{5}{2}$), adding an edge to already-large degrees yields a larger increase in MP and YP.*

This qualitative difference aligns with the strict hierarchy in Corollary 50.

6.4 A remark on structured extremals

Remark 56. *A common approach in extremal entropy questions is: (i) prove that a local transformation (often a 2-switch or a degree-shift) makes $\mathbf{p}^{(a)}$ more concentrated in the sense of majorization and hence decreases E_a by Schur-concavity; (ii) iterate until reaching a structured degree pattern; (iii) optimize within that family. For $a > 1$ (in particular $a = \frac{3}{2}, \frac{5}{2}$), the convexity of $t \mapsto t^a$ makes degree shifting toward higher-degree vertices especially effective (Remark 55). For $a = \frac{1}{2}$ (IP), the concavity of $t \mapsto t^{1/2}$ suggests a different sensitivity, while Corollary 51 permits transferring lower bounds proved for E_{YP} (or E_{MP}) to E_{IP}.*

ARTICLE IN PRESS

6.5 Illustrative molecular example: caffeine

To provide a molecular-scale illustration of the proposed descriptors, we consider the well-known molecule *caffeine* and compute both the classical degree-based entropies and the Prodeg entropies on its *hydrogen-suppressed* molecular graph (the same convention used throughout the QSPR experiments). Caffeine exhibits a non-uniform degree sequence (a mixture of degree-1, degree-2, and degree-3 vertices), so the induced degree-power probability vectors are non-uniform; increasing the exponent places relatively more mass on higher-degree vertices and therefore increases concentration.

Graph partition. The caffeine graph Figure 1 has the degree partition $n_1 = 5$ vertices of degree 1, $n_2 = 2$ vertices of degree 2, and $n_3 = 7$ vertices of degree 3 (total $n = 14$). This agrees with the edge partition since $|E| = 15$ and

$$\sum_{v \in V(G)} d(v) = 5 \cdot 1 + 2 \cdot 2 + 7 \cdot 3 = 30 = 2|E|.$$

For completeness, the edge partition Table 5 is $|E_{1,3}| = 5$, $|E_{2,2}| = 1$, $|E_{2,3}| = 2$, and $|E_{3,3}| = 7$, which also sums to $|E| = 15$.

Indices and entropy specialization. Let $k > 0$ denote the degree-power used for weighting, i.e., $w(v) = d(v)^k$, and define

$$S_k := \sum_{v \in V(G)} d(v)^k = 5 + 2 \cdot 2^k + 7 \cdot 3^k.$$

ARTICLE IN PRESS

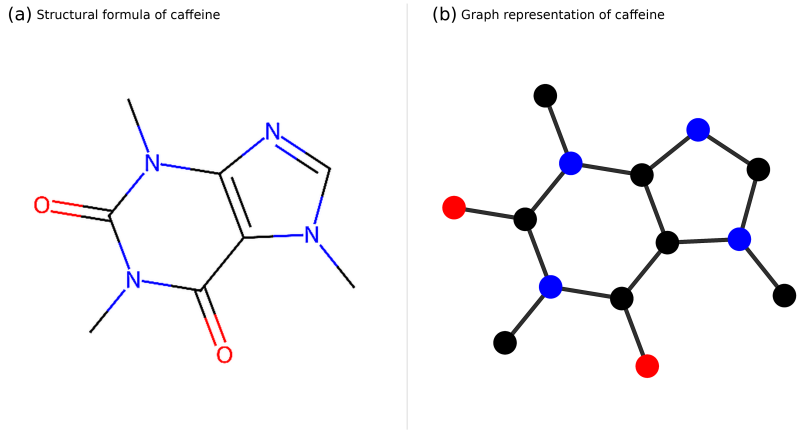

Figure 1: (a) Molecular structure and (b) hydrogen-suppressed molecular graph of caffeine.

Table 4: Vertex partition of the caffeine molecular graph by degree.

Vertex degree	1	2	3
Cardinality	5	2	7

Each degree class $d \in \{1, 2, 3\}$ induces the per-vertex probability $p(d) = d^k/S_k$. The corresponding Shannon entropy (natural logarithm) is

$$E_k(G) = -\sum_{v \in V(G)} p_v \log p_v = -\sum_{d \in \{1,2,3\}} n_d\, p(d)\, \log p(d)$$
$$= \log S_k - \frac{1}{S_k} \sum_{v \in V(G)} d(v)^k \log\big(d(v)^k\big) = \log S_k - \frac{k}{S_k}\Big(2 \cdot 2^k \log 2 + 7 \cdot 3^k \log 3\Big), \quad (26)$$

where the degree-1 term vanishes because $\log(1^k) = 0$.

Majorization hierarchy (interpretation). Consistent with the majorization/Schur-concavity principle established in Section 6.1, increasing k makes the probability vector more concentrated on high-degree vertices, hence decreases entropy. In particular, for caffeine the Prodeg family satisfies the strict hierarchy

$$E_{IP} = E_{1/2}(G) > E_{MP} = E_{3/2}(G) > E_{YP} = E_{5/2}(G),$$

reflecting the graph's degree heterogeneity (non-regularity). Table 6 reports the numerical values and compares the Prodeg entropies to the classical degree-based entropies computed from $k \in \{1, 2, 3, 4\}$.

ARTICLE IN PRESS

Table 5: Edge partition of the caffeine molecular graph by endpoint degrees.

Edge type	$E_{1,3}$	$E_{2,2}$	$E_{2,3}$	$E_{3,3}$
Cardinality	5	1	2	7

Table 6: Case study (caffeine): classical degree-based entropies and Prodeg entropies computed on the hydrogen-suppressed molecular graph (natural logarithm).

Molecule	E	E_M	E_F	E_Y	E_{IP}	E_{MP}	E_{YP}
Caffeine	2.539749	2.363424	2.222421	2.131429	2.610453	2.451220	2.286152

7 Relationship between degree-based entropies, information measures, and Bertz complexity

Beyond reporting descriptor values in Table 7, we quantify how two entropy families—(i) *classical degree-based entropies* and (ii) *Prodeg entropies*—track established information-theoretic and complexity measures, namely AvgIpc, $\log_{10}$(Ipc), and BertzCT, over the antibacterial compound set (C_1-C_{3219}, $N = 3219$) curated from the ChEMBL database. The goal is to determine whether the proposed graph-entropy descriptors behave consistently with widely used notions of molecular information content and structural complexity, and to assess whether Prodeg entropies provide signal beyond classical degree-entropies.

Analysis objectives. We (i) quantify associations between the *degree-entropy block* and the reference measures AvgIpc, $\log_{10}$(Ipc), and BertzCT; (ii) quantify associations between the *Prodeg-entropy block* and the same references; and (iii) compare the two entropy blocks to assess redundancy versus complementarity within the antibacterial chemical space.

ARTICLE IN PRESS

7.1 Descriptor blocks and analysis protocol

We consider three descriptor blocks (Table 7): (i) the **degree-entropy block** $\{E, E_M(C_i), E_F(C_i), E_Y(C_i)\}$, (ii) the **Prodeg-entropy block** $\{E_{IP}(C_i), E_{MP}(C_i), E_{YP}(C_i)\}$, and (iii) **reference measures** {AvgIpc, $\log_{10}$(Ipc), BertzCT}. All analyzed variables were available for the full ChEMBL-derived antibacterial set of $N = 3219$ compounds.

Associations were quantified using Pearson correlation (r, linear association) and Spearman correlation (ρ, rank-based monotone association). Multiple testing was controlled using the Benjamini-Hochberg false discovery rate (FDR) procedure[55]; given $N = 3219$ and the observed effect sizes, the reported associations remain significant after FDR. Because $\log_{10}$(Ipc) can exhibit extreme values, we interpret Pearson and Spearman jointly and use scatterplots to check for curvature and leverage effects.

7.2 Global structure and qualitative patterns

Figure 2 summarizes Pearson correlations among all entropy descriptors and reference measures. Two qualitative patterns are evident. First, the entropy descriptors (classical and

ARTICLE IN PRESS

Table 7: Entropy descriptors based on degree (Dehmer, first Zagreb, forgotten, and Yemen) and Prodeg entropy (inverse, misbalance, and Yemen), together with information measures (AvgIpc and $\log_{10}$(Ipc)) and the complexity index BertzCT, for the antibacterial compound set (C_1-C_{3219}). All values are rounded to two decimal places.

Compound	E	$E_M(C_i)$	$E_F(C_i)$	$E_Y(C_i)$	$E_{IP}(C_i)$	$E_{MP}(C_i)$	$E_{YP}(C_i)$	AvgIpc	BertzCT	$\log_{10}$(Ipc)
C_1	3.62	3.51	3.39	3.27	3.65	3.57	3.46	3.45	2758.20	13.43
C_2	2.33	2.17	2.00	1.83	2.38	2.25	2.08	2.32	581.40	4.23
C_3	3.41	3.25	3.10	2.98	3.47	3.33	3.17	3.32	1946.67	11.76
C_4	3.36	3.16	3.01	2.89	3.43	3.26	3.08	3.07	1970.18	11.20
C_5	2.42	2.27	2.10	1.93	2.47	2.35	2.19	2.32	544.19	3.80
C_6	3.30	3.22	3.10	2.96	3.32	3.26	3.16	3.37	2031.69	10.58
C_7	3.52	3.37	3.18	2.95	3.57	3.45	3.28	3.30	2556.72	14.97
C_8	2.66	2.42	2.15	1.90	2.74	2.55	2.28	2.83	929.08	5.61
C_9	3.07	2.94	2.82	2.72	3.12	3.01	2.88	2.93	1374.21	7.65
C_{10}	3.61	3.44	3.26	3.09	3.67	3.52	3.35	3.25	2737.49	12.62
C_{11}	3.49	3.34	3.14	2.92	3.54	3.42	3.25	3.16	2250.71	11.82
C_{12}	3.46	3.33	3.20	3.07	3.51	3.40	3.26	3.54	2265.09	12.47
C_{13}	3.14	2.98	2.84	2.73	3.20	3.06	2.91	2.94	1624.90	8.73
C_{14}	3.12	3.01	2.89	2.77	3.16	3.07	2.95	3.18	1582.33	8.58
C_{15}	2.98	2.92	2.81	2.68	2.99	2.95	2.87	3.25	1364.02	7.01
...	...	...	...	...	...	...	...	...	...	...
C_{3215}	2.87	2.72	2.59	2.49	2.92	2.80	2.65	3.00	992.62	6.64
C_{3216}	3.32	3.17	3.03	2.90	3.38	3.25	3.10	2.99	1899.07	10.02
C_{3217}	3.33	3.19	3.06	2.95	3.38	3.26	3.13	3.04	1855.86	9.16
C_{3218}	3.21	3.12	2.99	2.86	3.25	3.17	3.06	3.36	1560.09	7.88
C_{3219}	3.99	3.85	3.70	3.57	4.04	3.92	3.77	0.64	4193.80	44.77

Abbreviations used in Table 7: E = degree (Dehmer) entropy, E_M = first Zagreb entropy, E_F = forgotten entropy, E_Y = Yemen entropy, E_{IP} = inverse Prodeg entropy, E_{MP} = misbalance Prodeg entropy, E_{YP} = Yemen Prodeg entropy, AvgIpc = average information content, BertzCT = Bertz molecular complexity, $\log_{10}$(Ipc) = base-10 logarithm of the information content.

ARTICLE IN PRESS

Prodeg) form a tight cluster with very high mutual correlations, reflecting strong shared structural signal. Second, BertzCT and $\log_{10}$(Ipc) are strongly aligned with the entropy cluster, whereas AvgIpc shows weaker association and exhibits sign-inconsistency under Pearson correlation, suggesting that the AvgIpc-entropy relationship is not well captured by a globally linear model.

This geometry is supported by PCA (Figure 3), where PC1 explains 84.4% of the total variance and represents a dominant "complexity/entropy" axis shared by most variables, whereas PC2 (12.6%) captures secondary variation that primarily separates AvgIpc from the remaining descriptors.

7.3 Degree-based entropies vs. information and complexity measures

Table 8 reports Pearson and Spearman associations between the degree-entropy block and the reference measures. Three findings are noteworthy. First, degree-based entropies show a **strong positive association with BertzCT** (Pearson r = 0.839-0.922; Spearman

ARTICLE IN PRESS

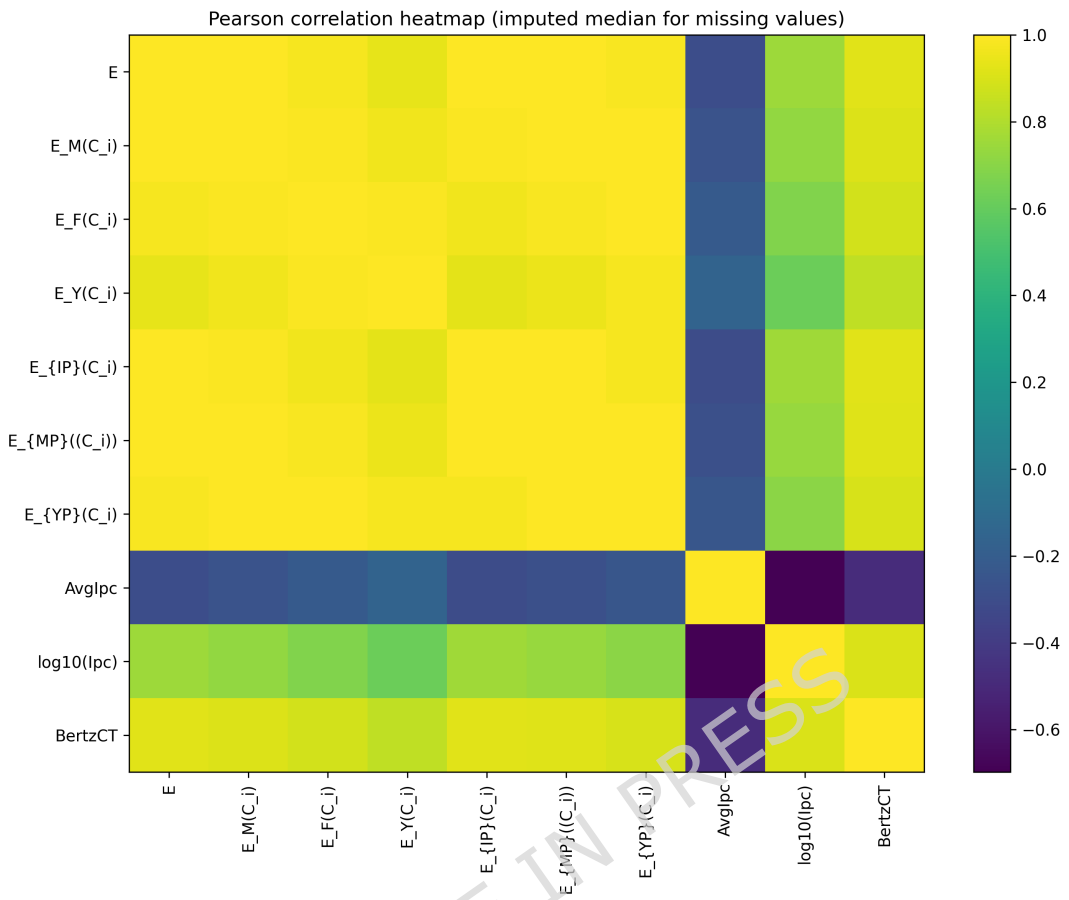

Figure 2: Pearson correlation heatmap for entropy descriptors (classical degree-based and Prodeg) and reference measures (AvgIpc, $\log_{10}$(Ipc), BertzCT), computed on the antibacterial set ($N = 3219$). Entropies form a tight cluster and show strong concordance with BertzCT and $\log_{10}$(Ipc), while AvgIpc exhibits weaker and sign-inconsistent linear association.

ARTICLE IN PRESS

$\rho = 0.945$-0.960), indicating that higher degree-entropy is strongly linked to increased molecular complexity as quantified by BertzCT. Second, they show a **moderate-to-strong positive association with** $\log_{10}$(Ipc) (Pearson $r = 0.621$-0.750; Spearman $\rho = 0.783$-0.811), consistent with an information-content interpretation. Third, the association with AvgIpc is **weaker and model-dependent**: Pearson correlations are negative (-0.159 to -0.294), whereas Spearman correlations are positive (0.236 to 0.311), consistent with a nonlinear yet broadly monotone relationship (i.e., curvature and/or leverage effects).

Figure 4 visualizes representative entropy-BertzCT relationships and shows an increasing trend with mild curvature, explaining why Spearman ρ is slightly larger than Pearson r.

At the block level, the mean absolute correlation with the three reference measures is $\mathcal{A}_{\text{deg}} = 0.605$ (Pearson) and $\mathcal{A}_{\text{deg}} = 0.700$ (Spearman), indicating strong overall alignment dominated by BertzCT and $\log_{10}$(Ipc).

ARTICLE IN PRESS

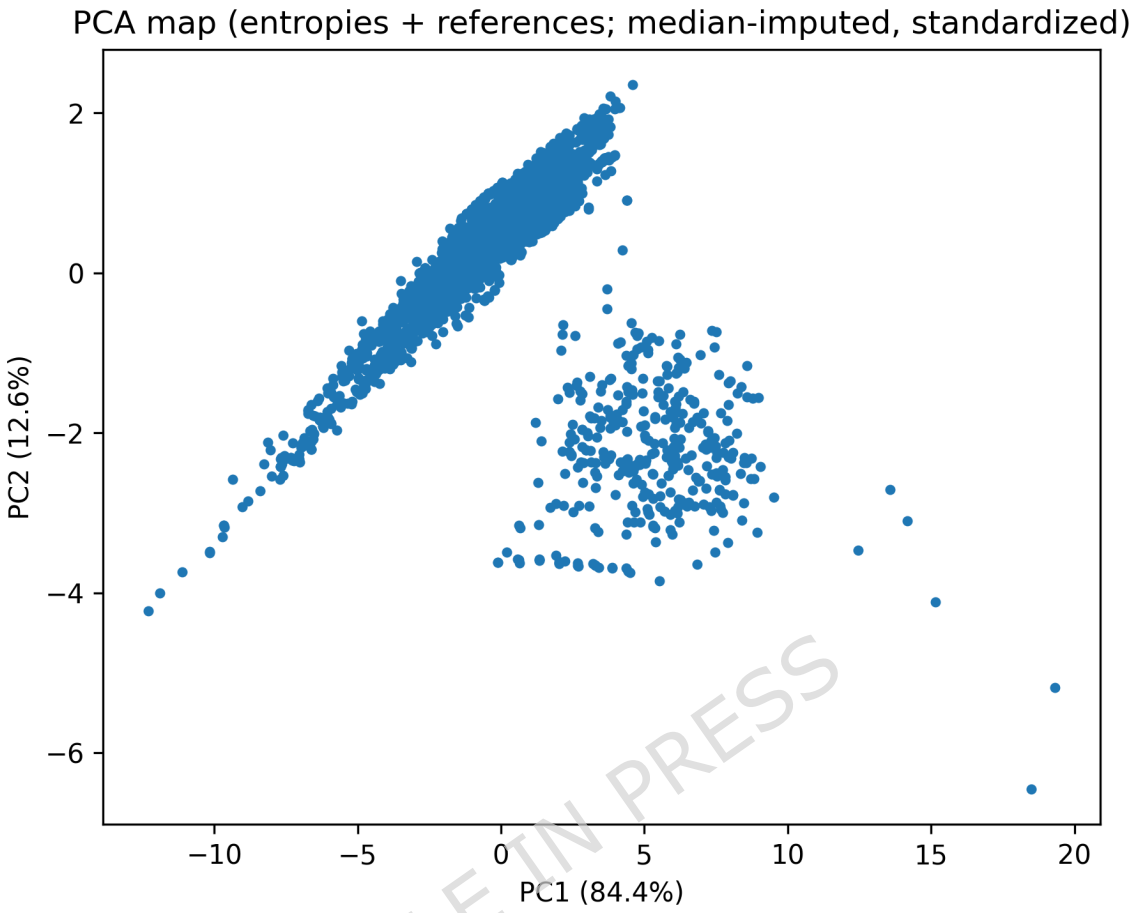

Figure 3: PCA map of standardized descriptors (entropies + reference measures). PC1 explains 84.4% of the variance and captures the dominant shared complexity/entropy axis; PC2 (12.6%) captures secondary variation that separates AvgIpc from the other measures.

7.4 Prodeg entropies vs. information and complexity measures

Table 9 reports the same analysis for the Prodeg-entropy block. Prodeg entropies show a **strong positive association with BertzCT** (Pearson $r = 0.899$-0.922; Spearman $\rho = 0.960$-0.972), and a **moderate-to-strong association with** $\log_{10}$(Ipc) (Pearson $r = 0.703$-0.755; Spearman $\rho = 0.794$-0.815). The AvgIpc relationship remains weaker and exhibits the same Pearson/Spearman sign discrepancy (negative Pearson, positive Spearman), again consistent with nonlinearity rather than an absence of monotone association.

At the block level, Prodeg entropies show slightly higher alignment with the reference measures: $\mathcal{A}_{\text{prodeg}} = 0.640$ (Pearson) and $\mathcal{A}_{\text{prodeg}} = 0.711$ (Spearman). Figure 5 visualizes representative Prodeg entropy-BertzCT relationships.

ARTICLE IN PRESS

Table 8: Associations between classical degree-based entropies and reference measures (AvgIpc, $\log_{10}$(Ipc), BertzCT) for $N = 3219$ antibacterial compounds. Pearson r measures linear association; Spearman ρ measures monotone association.

Descriptor	r(AvgIpc)	ρ(AvgIpc)	$r(\log_{10}\mathrm{Ipc})$	$\rho(\log_{10}\mathrm{Ipc})$	r(BertzCT)	ρ(BertzCT)
E	-0.294	0.311	0.750	0.811	0.922	0.960
$E_M(C_i)$	-0.263	0.297	0.722	0.799	0.910	0.957
$E_F(C_i)$	-0.217	0.267	0.680	0.789	0.885	0.952
$E_Y(C_i)$	-0.159	0.236	0.621	0.783	0.839	0.945

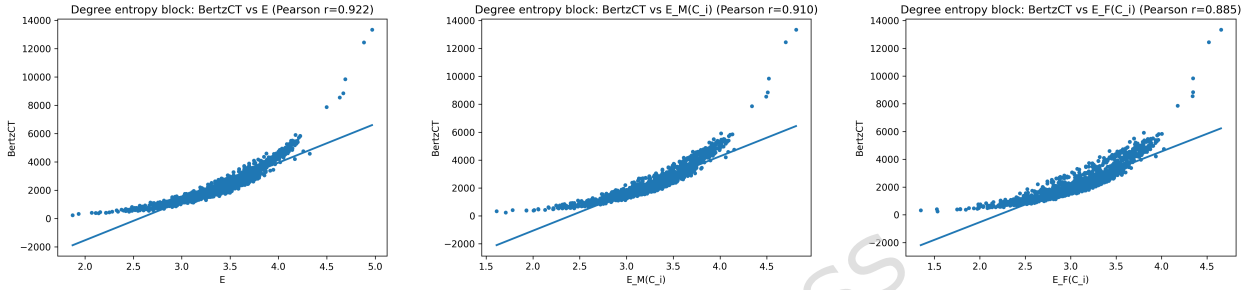

Figure 4: Scatterplots of BertzCT vs. selected classical degree-based entropies for $N = 3219$. The relationships are strongly increasing with mild curvature, consistent with high Spearman correlations and slightly lower Pearson correlations.

7.5 Comparing degree-entropies and Prodeg entropies: redundancy vs. complementarity

Cross-block correlations (Table 10) are extremely high (Pearson $r \approx 0.93$-0.999), indicating that, over this antibacterial chemical space, Prodeg entropies largely track the same dominant structural signal as classical degree-based entropies.

To test whether Prodeg entropies add incremental signal for the reference measures, we fit two models for each target $R \in \{\mathrm{AvgIpc}, \log_{10}(\mathrm{Ipc}), \mathrm{BertzCT}\}$:

$$R \sim \{E, E_M, E_F, E_Y\}, \qquad R \sim \{E, E_M, E_F, E_Y, E_{IP}, E_{MP}, E_{YP}\}.$$

Table 11 reports adjusted R^2 and cross-validated (CV) performance. Adding the Prodeg block yields consistent but modest gains: $\Delta\mathrm{adj}R^2 \approx 0.048$ for AvgIpc, ≈ 0.030 for $\log_{10}$(Ipc), and ≈ 0.009 for BertzCT, with corresponding reductions in CV-RMSE. Thus, within this ChEMBL-derived antibacterial chemical space, the Prodeg entropies act mainly as compact complementary descriptors rather than as a fully independent source of structural information.

Multicollinearity diagnostics (Table 12) reveal extremely large VIF values (up to $\sim 10^9$), which is expected given the near-duplicate nature of the entropy families. Consequently, while the incremental gains are reproducible numerically, the combined unregularized linear model is not suitable for stable coefficient inference. For downstream predictive modeling, this motivates either selecting a compact representative entropy subset (one per highly cor-

ARTICLE IN PRESS

Table 9: Associations between Prodeg entropies and reference measures for $N = 3219$ antibacterial compounds.

Descriptor	r(AvgIpc)	ρ(AvgIpc)	$r(\log_{10}\mathrm{Ipc})$	$\rho(\log_{10}\mathrm{Ipc})$	r(BertzCT)	ρ(BertzCT)
$E_{IP}(C_i)$	-0.301	0.305	0.755	0.815	0.922	0.972
$E_{MP}(C_i)$	-0.280	0.284	0.738	0.806	0.918	0.968
$E_{YP}(C_i)$	-0.242	0.250	0.703	0.794	0.899	0.960

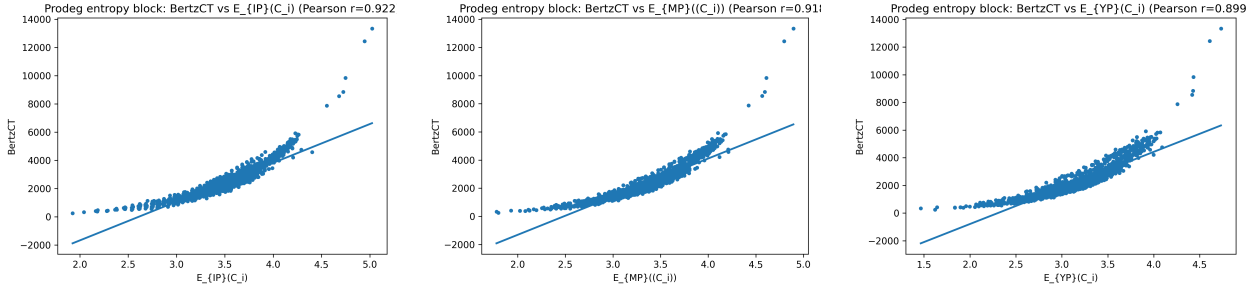

Figure 5: Scatterplots of BertzCT vs. Prodeg entropies for $N = 3219$, showing strong monotone increases and supporting their close alignment with molecular complexity.

Table 10: Cross-block Pearson correlations between classical degree-entropies (rows) and Prodeg entropies (columns), $N = 3219$. Near-unity correlations indicate strong redundancy between entropy families in this dataset.

	$E_{IP}(C_i)$	$E_{MP}(C_i)$	$E_{YP}(C_i)$
E	0.999	0.998	0.986
$E_M(C_i)$	0.987	0.998	0.998
$E_F(C_i)$	0.966	0.985	0.998
$E_Y(C_i)$	0.929	0.951	0.976

related cluster) or using regularized regression (ridge/elastic net) or dimension reduction (PCA).

7.6 Quality-control note on extreme values in $\log_{10}$(Ipc)

The distribution of $\log_{10}$(Ipc) contains extreme high-end values (flagged by a 3×IQR rule in a small subset of compounds). Because such values can inflate leverage in least-squares fits, we relied on Spearman ρ and visual inspection to verify that the dominant trends (strong monotone increases of BertzCT and $\log_{10}$(Ipc) with entropy) remain stable. The close agreement between Pearson and Spearman for the strongest relationships (particularly with BertzCT) supports the robustness of the conclusions.

ARTICLE IN PRESS

Table 11: Incremental linear modeling of each reference measure using (Model 1) the degree-entropy block only and (Model 2) degree + Prodeg blocks. CV metrics are reported as mean over folds.

Target	adjR^2 (M1)	adjR^2 (M2)	ΔadjR^2	CV-R^2 (M1)	CV-R^2 (M2)	ΔCV-R^2	CV-RMSE (M1)	CV-RMSE (M2)	ΔRMSE
AvgIpc	0.204	0.252	0.048	0.195	0.238	0.043	0.668	0.649	-0.018
$\log_{10}$(Ipc)	0.621	0.651	0.030	0.618	0.647	0.029	5.811	5.586	-0.225
BertzCT	0.856	0.865	0.009	0.857	0.866	0.009	356.127	344.309	-11.817

Table 12: Variance inflation factors (VIF) for the full entropy predictor set (degree + Prodeg). Extremely large VIF values indicate severe multicollinearity due to near-duplicate descriptors.

Variable	VIF
$E_M(C_i)$	2.35e+09
$E_{MP}(C_i)$	1.88e+09
$E_{YP}(C_i)$	9.72e+08
E	5.14e+08
$E_F(C_i)$	9.17e+07
$E_{IP}(C_i)$	3.13e+07
$E_Y(C_i)$	3.12e+05

ARTICLE IN PRESS

7.7 Implications for descriptor design and modeling

Across the antibacterial chemical space, both entropy families act as strong proxies for established complexity measures: BertzCT exhibits the strongest concordance, followed by $\log_{10}$(Ipc). Prodeg entropies show slightly higher block-level alignment with the reference measures, but cross-block correlations indicate that the two entropy blocks are largely redundant. Therefore, for predictive modeling it is preferable to use a compact entropy set (or PCA/regularization) to mitigate multicollinearity while preserving the dominant complexity/entropy signal.

ARTICLE IN PRESS

8 QSPR modeling of physicochemical properties using entropy descriptors

8.1 Overall modelling strategy

To evaluate the practical usefulness of the proposed Prodeg-entropy family in quantitative structure-property relationships (QSPR), we designed a controlled benchmark in which *only* degree-entropy descriptors extracted from the molecular graph were used as predictors. The central question is whether the Prodeg entropies provide (i) *competitive* predictive signal relative to widely used classical degree-based entropies and/or (ii) *complementary* information when both families are fused.

We considered nine continuous physicochemical properties that jointly characterize molecular size, polarity, hydrogen-bonding capacity, flexibility, and drug-likeness. Each endpoint was modeled independently under the same cross-validation protocol and evaluation criteria, enabling a direct, block-wise comparison of descriptor sets. Three descriptor blocks were assessed: (i) a *classical* degree-entropy block, (ii) a *Prodeg* entropy block, and (iii) a *combined* block concatenating all entropies. For each endpoint and descriptor block, we benchmarked a representative set of regression algorithms spanning linear and nonlinear hypothesis classes, and we report performance from out-of-fold (OOF) predictions to mitigate optimistic bias.

For the block-wise comparisons, the "best model" for a given endpoint–descriptor-block pair was selected independently as the candidate regressor with the smallest out-of-fold RMSE within that block. Therefore, different descriptor blocks may be represented by different algorithms for the same endpoint. This rule is useful for comparing the best predictive performance attainable from each descriptor block under the same candidate model set, but the resulting cross-block differences should be interpreted as block-plus-model performance differences rather than as algorithm-controlled comparisons with a fixed learner.

ARTICLE IN PRESS

8.2 Dataset and target properties

All compounds were obtained from the ChEMBL database, a manually curated and FAIR repository of bioactive molecules with drug-like properties [54]. Since identical chemical structures may appear multiple times in ChEMBL (e.g., linked to different assays or experimental settings), structural redundancy was removed using a stringent uniqueness criterion based on the pair (Molecule ChEMBL ID, IUPAC InChIKey). After de-duplication, only ChEMBL molecules with available entropy descriptors and reported physicochemical properties were retained for modeling.

It should be emphasized that the empirical QSPR component is based exclusively on a ChEMBL-derived antibacterial compound set. Hence, the predictive conclusions should be interpreted primarily within this antibacterial chemical space. The mathematical definitions and closed-form entropy results are graph-theoretic and apply to arbitrary connected graphs, but the observed QSPR ranking of descriptor blocks may change for other disease areas, assay types, or broader chemical libraries. External validation on additional therapeutic classes and more structurally diverse compound collections is therefore required before making general claims about performance across chemical space.

ARTICLE IN PRESS

The nine modeled targets are standard physicochemical descriptors reported for the antibacterial compound set (C_1-C_{3219}) (Table 13): MolLogP, MolMR, Molecular Weight, HBA, HBD, #Rotatable Bonds, QED Weighted, AlogP, and Polar Surface Area. Endpoint-specific sample sizes are reported together with model performance in Tables 14-15. The variation in N across endpoints reflects endpoint-wise complete-case filtering after curation: some compounds had valid molecular structures and entropy descriptors but lacked a valid value for one or more physicochemical targets. Target values were not imputed; instead, each endpoint was modeled using the largest subset with complete response and predictor information.

Data set and modeling protocol

Data sources and curation. The compound collection (C_1-C_{3219}) was assembled exclusively from the ChEMBL database and curated to remove duplicates and inconsistent records. Molecular structures were standardized (e.g., salt removal and charge normalization when applicable) and represented as canonical SMILES strings. All target properties (Table 13) and entropy-based predictors (Table 7) were computed consistently for the full set. After endpoint-wise filtering, MolMR and Molecular Weight were available for all 3219 compounds, HBA, HBD, rotatable bonds, QED, AlogP, and PSA were available for 3213 compounds, and MolLogP was available for 3062 compounds. These differences are due to missing or invalid endpoint records after standardization and curation, not to different descriptor-calculation procedures.

Performance metrics. Predictive performance was quantified by the coefficient of determination (R^2), root-mean-square error (RMSE), and mean absolute error (MAE):

$$R^2 = 1 - \frac{\sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}{\sum_{i=1}^{n} (y_i - \overline{y})^2}, \qquad RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \widehat{y}_i)^2}, \qquad MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \widehat{y}_i| , \tag{27}$$

where y_i and $\widehat{y}_i$ denote the observed and predicted responses and $\overline{y}$ is the sample mean. For k-fold CV, the dataset was partitioned into k disjoint folds; each fold served once for validation while the remaining $k-1$ folds were used for fitting, and performance was computed from out-of-fold (OOF) predictions.

Entropy descriptor blocks and QSPR evaluation

To quantify the predictive utility of the proposed Prodeg entropies in a QSPR setting, we modeled nine standard physicochemical endpoints reported for the antibacterial compound set (C_1-C_{3219}) (Table 13): MolLogP, MolMR, Molecular Weight, HBA, HBD, #Rotatable Bonds, QED Weighted, AlogP, and Polar Surface Area. Each molecule was represented only by degree-entropy descriptors derived from its molecular graph, grouped into three blocks: (i) a *classical* degree-entropy block consisting of the degree entropy E (Dehmer), the first Zagreb entropy E_M, the forgotten entropy E_F, and the Yemen entropy E_Y; (ii) a *Prodeg* block consisting of the inverse Prodeg entropy E_{IP}, the misbalance Prodeg entropy E_{MP}, and the Yemen Prodeg entropy E_{YP}; and (iii) the *combined* block that concatenates all

ARTICLE IN PRESS

Table 13: Physicochemical descriptors for the compound set (C_1-C_{3219}).

No	Molecular Formula	MolLogP	MolMR	MW	HBA	HBD	RB	QED	AlogP	PSA
1.	$C_{28}H_{28}F_2N_4O_5$	3.50	141.03	538.55	8.00	2.00	9.00	0.40	3.50	105.39
2.	$C_8H_{10}N_2S$	1.28	49.50	166.25	2.00	1.00	2.00	0.67	1.28	38.91
3.	$C_{21}H_{29}N_2NaO_9S$	-3.42	113.02	508.53	9.00	3.00	9.00	0.24	0.91	151.70
4.	$C_{22}H_{24}N_2O_8$	-0.35	109.77	444.44	9.00	6.00	2.00	0.33	-0.35	181.62
5.	$C_8H_6INO_2$	2.54	55.19	275.05	2.00	0.00	2.00	0.47	2.54	43.14
6.	$C_{22}H_{24}N_6$	4.63	113.25	372.48	6.00	2.00	7.00	0.49	4.63	67.66
7.	$C_{28}H_{44}N_4O_4$	3.34	138.89	500.68	5.00	3.00	11.00	0.25	3.34	101.98
8.	$C_8H_{13}N_3O_4S$	0.53	57.95	247.28	6.00	0.00	5.00	0.56	0.53	95.10
9.	$C_{17}H_{14}BrNO_4$	3.66	87.88	376.21	4.00	1.00	1.00	0.83	3.66	59.00
10.	$C_{26}H_{23}F_3N_4O_6S$	3.28	139.97	576.55	6.00	2.00	3.00	0.69	2.04	101.45
11.	$C_{27}H_{23}N_3O_5$	1.89	129.91	469.50	6.00	6.00	5.00	0.15	1.89	144.91
12.	$C_{22}H_{29}N_7O_5$	-0.79	123.26	471.52	11.00	4.00	8.00	0.33	-0.79	160.88
13.	$C_{17}H_{19}F_2N_3O_3$	1.78	90.13	351.35	5.00	2.00	4.00	0.88	1.78	88.56
14.	$C_{17}H_{18}FN_3O_3$	1.58	88.48	331.35	5.00	2.00	3.00	0.89	1.58	74.57
15.	$C_{16}H_{12}N_2O_2$	3.46	77.59	264.28	3.00	1.00	2.00	0.77	3.46	47.14
...	...	...	...	...	...	...	...	...	...	...
3215.	$C_{12}H_{14}NNaO_5S$	-3.96	64.40	307.30	5.00	2.00	3.00	0.73	0.37	87.07
3216.	$C_{23}H_{20}Br_2ClNO_3$	8.05	128.25	553.68	3.00	2.00	5.00	0.34	8.05	58.56
3217.	$C_{20}H_{14}F_3N_3O_4$	3.15	102.43	417.34	6.00	2.00	3.00	0.50	3.15	95.13
3218.	$C_{17}H_{10}N_4O_3S_2$	4.10	103.52	382.43	7.00	1.00	3.00	0.42	4.10	97.49
3219.	$C_{36}H_{74}N_{10}O_{12}$	-7.90	209.07	839.05	22.00	16.00	18.00	0.06	-7.90	408.98

Abbreviations used in Table 13: MolLogP = molecular logP, MolMR = molecular refractivity ($cm^3\,mol^{-1}$), MW = molecular weight ($g\,mol^{-1}$), HBA = hydrogen bond acceptors (count), HBD = hydrogen bond donors (count), RB = #Rotatable Bonds (count), QED = quantitative estimate of drug-likeness, AlogP = atomic fragment-based logP, PSA = polar surface area ($Å^2$).

ARTICLE IN PRESS

seven entropies. This design enables a direct assessment of whether Prodeg entropies provide information that is (a) competitive with classical degree-entropies using fewer descriptors, and/or (b) complementary when fused with them.

Modeling protocol. For each endpoint, we retained all compounds with available target values and complete entropy descriptors, leading to endpoint-dependent sample sizes (N = 3062-3219; Tables 14-15). All models were evaluated with 5-fold cross-validation (KFold; shuffled; random seed 42), with all randomized steps controlled by a fixed seed: the fold assignment used **random_state=42**, and stochastic estimators (e.g., Random Forest, Extra Trees, and histogram gradient boosting) were initialized with the same setting when applicable. The full pipeline, including preprocessing within each training fold, is provided with the accompanying code to support exact replication. Performance was computed from out-of-fold (OOF) predictions. We report R^2, RMSE, and MAE to balance goodness-of-fit with scale-aware error measures. As candidate regressors, we considered a representative model zoo spanning linear and nonlinear hypothesis classes: (i) linear regression and regularized linear models (Ridge, ElasticNet) with standardization, and (ii) tree-based ensembles

ARTICLE IN PRESS

(Random Forest, Extra Trees) and histogram gradient boosting, which can capture nonlinearities and descriptor interactions. Missing values (when present) were handled via median imputation inside each training fold. This evaluation design was intentionally constrained (seven scalar predictors at most) to isolate the contribution of entropy descriptors without confounding from high-dimensional fingerprints or atom-typed fragments.

Overall predictive performance. Table 14 reports the best-performing model per endpoint (minimum RMSE across all models and descriptor blocks). Tree ensembles dominated the best models, indicating that entropy-property relationships are generally nonlinear and benefit from interaction modeling. The highest predictability was observed for MolMR ($R^2 = 0.970$) and Molecular Weight ($R^2 = 0.945$), consistent with the fact that degree-entropies encode global size/branching characteristics. Good predictive performance was also obtained for QED Weighted ($R^2 = 0.714$), #Rotatable Bonds ($R^2 = 0.677$), and Polar Surface Area ($R^2 = 0.660$). Moderate performance was found for HBA ($R^2 = 0.648$) and AlogP ($R^2 = 0.429$), while MolLogP was the most challenging endpoint ($R^2 = 0.268$). This behavior is expected because lipophilicity depends strongly on atom-level chemistry (functional groups and heteroatom patterns) that is only partially reflected by degree-based graph summaries. To improve MolLogP prediction within the same learning framework, a natural next step is to use hybrid descriptor blocks that combine the proposed Prodeg entropies (capturing global size/branching/topological dispersion) with complementary chemistry-sensitive features that encode local functionality. Examples include simple fragment- and composition-based descriptors (heteroatom counts, halogen counts, aromatic atom fraction, ring counts), hydrogen-bonding related counts, and/or standard substructure fingerprints (e.g., circular fingerprints) that better reflect functional-group patterns driving lipophilicity. Feature engineering that couples global topology to local chemistry (e.g., interactions such as entropy $\times$ heteroatom fraction or entropy $\times$ aromatic fraction) is also expected to be beneficial, especially for tree-ensemble models that can exploit nonlinear interactions.

Classical vs. Prodeg entropies. A block-wise comparison is provided in Table 15. Using only three Prodeg descriptors, the Prodeg block achieved competitive performance for size-related endpoints, e.g., MolMR ($R^2 = 0.968$) and Molecular Weight ($R^2 = 0.941$), despite having fewer variables than the classical block. Across the full endpoint set, the Prodeg block alone did not outperform the classical block in RMSE, but the performance gap was typically small for endpoints driven primarily by global structural characteristics. Importantly, when Prodeg and classical entropies were combined, the *combined* block achieved the lowest RMSE for 6 of 9 endpoints (MolLogP, HBA, HBD, QED Weighted, AlogP, Polar Surface Area), with RMSE reductions of about 0.3-1.0% relative to the best classical-only models (Table 15). Although modest in magnitude, these gains are consistent across diverse endpoints and support the interpretation that Prodeg entropies provide compact complementary information beyond classical degree-based entropy measures.

Visualization and diagnostics. Figure 6 summarizes the best R^2 achieved per endpoint. In addition, for each endpoint we provide observed-vs-predicted OOF scatter plots for the corresponding best model (Fig. 7), which serve as a diagnostic for systematic bias and het-

ARTICLE IN PRESS

Table 14: Best-performing QSPR model for each physicochemical endpoint based on 5-fold cross-validation (OOF predictions). For each endpoint, the selected row corresponds to the minimum RMSE across all evaluated descriptor blocks and candidate regressors; hence both the descriptor set and algorithm may differ between endpoints.

Endpoint	N	Descriptor set	Model	R^2	RMSE	MAE
MolLogP	3062	Combined	RF	0.268	2.217	1.552
MolMR	3219	Classical	RF	**0.970**	7.243	5.218
Molecular Weight	3219	Classical	RF	0.**945**	37.486	23.466
HBA	3213	Combined	RF	0.648	1.977	1.399
HBD	3213	Combined	HGB	0.475	1.426	0.974
#Rotatable Bonds	3213	Classical	ET	0.677	2.898	1.579
QED Weighted	3213	Combined	RF	0.714	0.136	0.101
AlogP	3213	Combined	HGB	0.429	1.905	1.385
Polar Surface Area	3213	Combined	RF	0.660	32.457	23.203

eroscedasticity. A compact summary of the effect of adding Prodeg entropies on RMSE (relative to the classical block) is shown in Fig. 8.

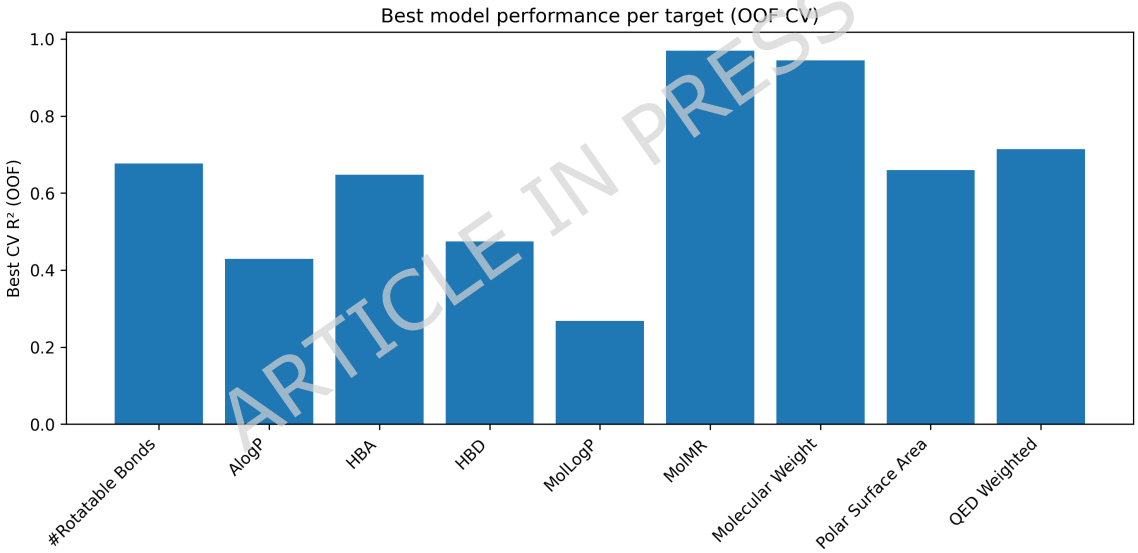

Figure 6: Best cross-validated performance (R^2 from OOF predictions) achieved for each physicochemical endpoint across all evaluated models and descriptor blocks.

Limitations and outlook. This experiment is intentionally stringent: each molecule is represented by only seven scalar degree-entropy descriptors. While this setting is appropriate for isolating the value of the proposed Prodeg entropies, higher accuracy can be expected by augmenting the feature space with atom-type-aware descriptors (e.g., fingerprints or fragment counts) and by exploring multi-task learning across correlated endpoints. Future work should also assess the stability of the observed RMSE gains under stricter validation protocols, such as scaffold-based splits, repeated cross-validation, or bootstrap resampling, and investigate model interpretability (e.g., feature importance or SHAP analyses) to clarify

ARTICLE IN PRESS

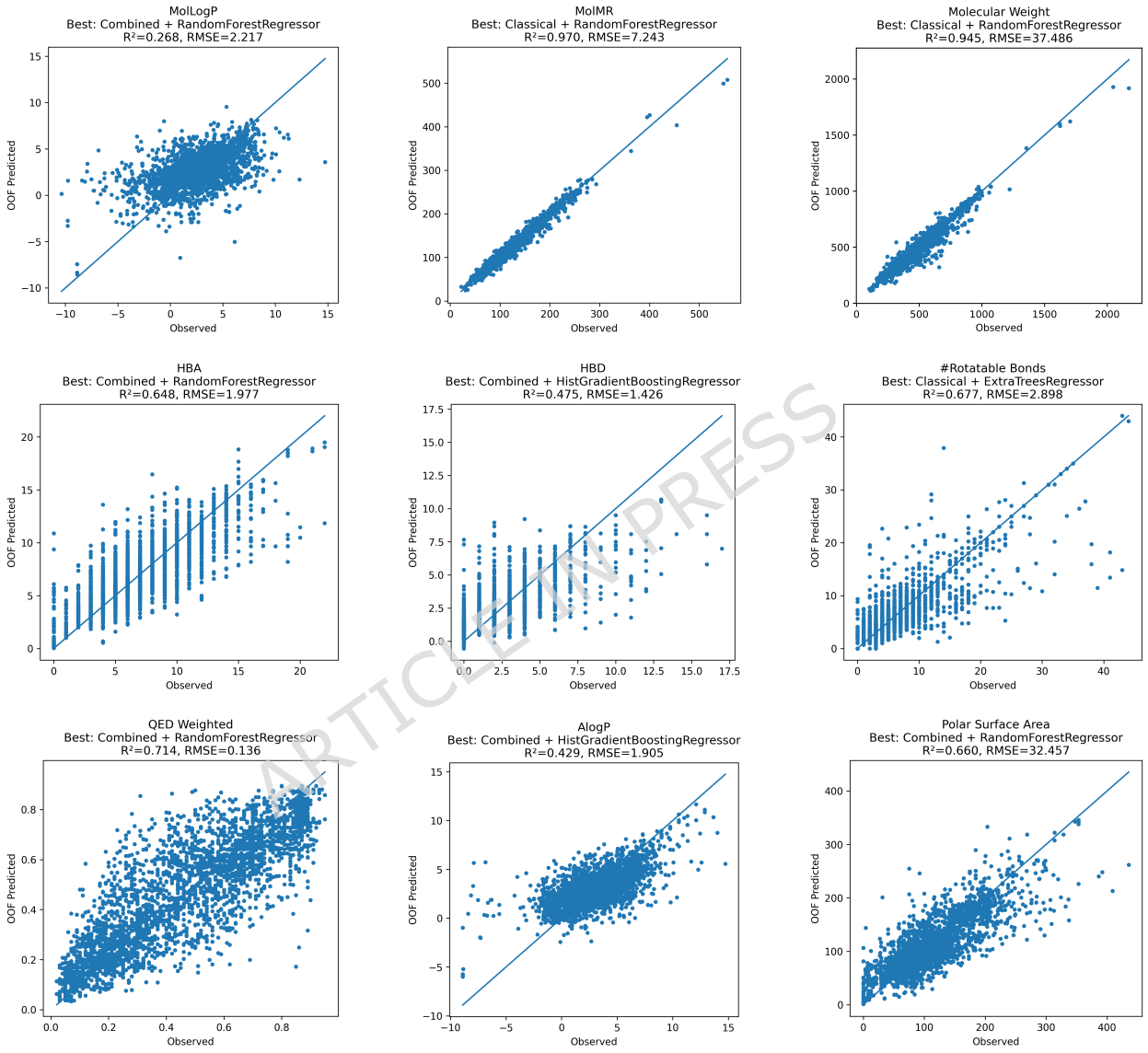

Figure 7: Observed vs. predicted values from out-of-fold (OOF) predictions for the best model of each endpoint. The diagonal line indicates perfect agreement.

ARTICLE IN PRESS

Table 15: Cross-validated QSPR performance for predicting physicochemical properties using (i) classical degree-based entropies, (ii) Prodeg entropies, and (iii) their combination. For each endpoint and descriptor block, the listed model is the independently selected best regressor, defined by the minimum out-of-fold RMSE among the evaluated algorithms. Bold RMSE indicates the winning descriptor block for that endpoint.

		Classical entropies (E, E_M, E_F, E_Y)			Prodeg entropies (E_{IP}, E_{MP}, E_{YP})			Combined (all entropies)				
Endpoint	N	Model	R^2	RMSE	Model	R^2	RMSE	Model	R^2	RMSE	Winner	ΔRMSE(%)
MolLogP	3062	HGB	0.258	2.233	HGB	0.235	2.267	RF	0.268	2.217	Combined	0.716
MolMR	3219	RF	0.970	7.243	RF	0.968	7.490	RF	0.970	7.266	Classical	-0.315
Molecular Weight	3219	RF	0.945	37.486	RF	0.941	38.967	RF	0.944	37.849	Classical	-0.969
HBA	3213	ET	0.643	1.990	RF	0.597	2.116	RF	0.648	1.977	Combined	0.665
HBD	3213	HGB	0.465	1.439	HGB	0.433	1.481	HGB	0.475	1.426	Combined	0.942
#Rotatable Bonds	3213	ET	0.677	2.898	ET	0.569	3.346	ET	0.668	2.940	Classical	-1.427
QED Weighted	3213	RF	0.712	0.137	RF	0.702	0.139	RF	0.714	0.136	Combined	0.501
AlogP	3213	RF	0.423	1.916	RF	0.356	2.024	HGB	0.429	1.905	Combined	0.578
Polar Surface Area	3213	RF	0.653	32.794	RF	0.613	34.642	RF	0.660	32.457	Combined	1.026

Notes: R^2, RMSE, and MAE are computed from out-of-fold (OOF) predictions under 5-fold cross-validation. ΔRMSE(%) is the percentage RMSE improvement of the Combined block relative to the Classical block (positive values indicate improvement). Model abbreviations: RF = Random Forest, ET = Extra Trees, HGB = Histogram Gradient Boosting. Because model selection is performed independently within each endpoint–descriptor-block pair, different blocks may use different algorithms. Therefore, cross-block comparisons summarize the best attainable performance of each descriptor block under the common candidate model set; they should not be interpreted as comparisons under a single fixed regression algorithm. Endpoint-specific N values reflect complete-case filtering for the corresponding target variable.

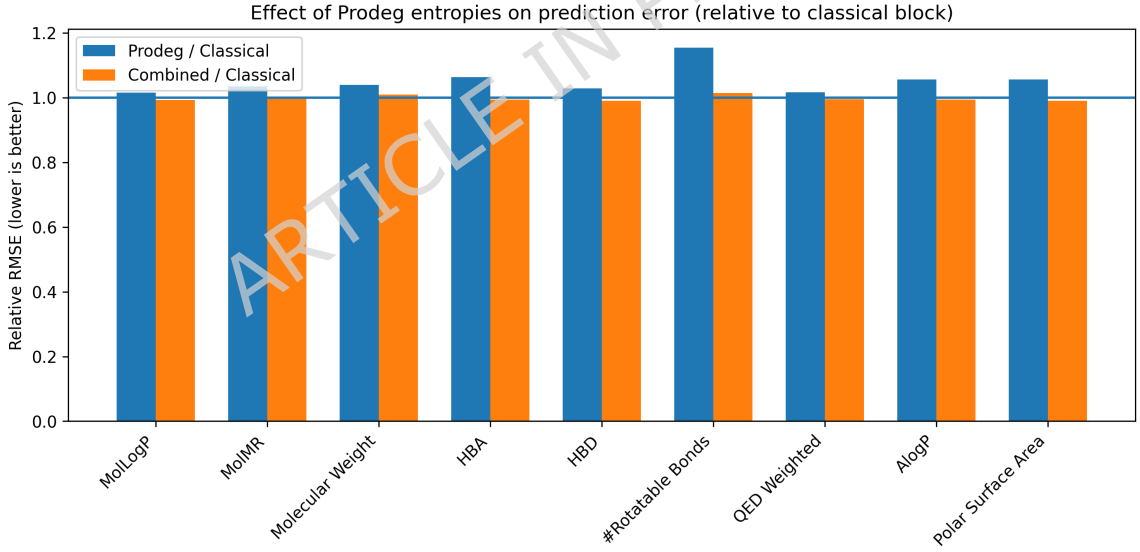

Figure 8: Relative RMSE (lower is better) of the Prodeg-only and Combined descriptor blocks normalized by the best classical-only RMSE for each endpoint. Values below 1 indicate improved prediction compared with the classical block.

which entropy components drive improvements. A further limitation is that the data set is restricted to antibacterial compounds curated from ChEMBL. Consequently, the empirical findings should not be overgeneralized to unrelated disease types, assay contexts, or highly different regions of chemical space without external validation. Nevertheless, the

ARTICLE IN PRESS

results here demonstrate that, within this curated antibacterial benchmark, degree-based entropy descriptors alone already recover substantial variance for several key properties and that Prodeg entropies contribute complementary signal when fused with classical entropy measures.

ARTICLE IN PRESS

ARTICLE IN PRESS

Conclusion and future directions

This paper develops a unified theoretical and applied framework for three Prodeg-type degree-power invariants—the Inverse Prodeg index $IP(G) = \sum_{v\in V(G)} d_G(v)^{1/2}$, the Misbalance Prodeg index $MP(G) = \sum_{v\in V(G)} d_G(v)^{3/2}$, and the Yemen Prodeg index $YP(G) = \sum_{v\in V(G)} d_G(v)^{5/2}$—together with their associated Shannon entropies $E_{IP}(G)$, $E_{MP}(G)$, and $E_{YP}(G)$. On the theoretical side, we derived explicit closed-form expressions for representative graph families and established sharp extremal behavior. In particular, for each of the three Prodeg entropies, the universal upper bound $\log n$ is attained if and only if the graph is regular. We also obtained bounds in terms of degree extremes and concentration via the heaviest probability weight, proved unified edge-addition increment identities, and established a tensor-product principle: for $a > 0$, the degree power sum $S_a(G) = \sum_{v\in V(G)} d_G(v)^a$ is multiplicative under the tensor (Kronecker) product, which yields additivity of $E_a(G)$ and the Nordhaus–Gaddum-type inequality $E_a(G) + E_a(\overline{G}) \leqslant 2\log n$ whenever both entropies are defined. Finally, using majorization, we proved the monotone exponent hierarchy

$$E_{IP}(G) \geqslant E_{MP}(G) \geqslant E_{YP}(G),$$

with equality throughout if and only if G is regular. Thus, increasing the exponent increases the concentration of the induced degree-power distribution and decreases entropy.

Beyond theory, we examined the descriptor behavior of Prodeg entropies over the antibacterial compound set (C_1-C_{3219}, $N = 3219$) curated from the ChEMBL database. Correlation and PCA analyses show that both classical degree-entropies and Prodeg entropies align strongly with established molecular complexity measures, especially BertzCT and $\log_{10}$(Ipc), whereas the association with AvgIpc is weaker and nonlinear. Cross-block correlations indicate substantial redundancy between the classical and Prodeg entropy families in this chemical space; however, incremental analyses show that adding Prodeg entropies still yields consistent, though modest, gains.

We also assessed the practical utility of the entropy descriptors in a controlled QSPR benchmark using only entropy-based predictors for nine physicochemical endpoints. Tree-ensemble models provided the best overall performance, with especially high predictability for size-related endpoints such as MolMR ($R^2 = 0.970$) and Molecular Weight ($R^2 = 0.945$). The Prodeg block remained competitive despite using fewer variables than the classical block, and the combined block achieved the lowest RMSE for most endpoints, supporting the view that Prodeg entropies provide compact complementary information even in an entropy-only setting. In contrast, MolLogP remained the most challenging endpoint ($R^2 = 0.268$), which is consistent with lipophilicity depending on chemistry-specific local features not fully captured by degree-based graph summaries.

Several directions naturally follow from the present results. From a theoretical perspective, it would be interesting to obtain complete extremal characterizations of E_{MP} and E_{YP} within constrained graph families, to formalize stability properties of E_a under local perturbations, and to investigate related entropy identities under other graph operations, especially the Cartesian product and related product-type constructions. From an applied perspective, future QSPR/QSAR studies should integrate Prodeg entropies with chemistry-sensitive descriptors, adopt stricter validation protocols such as scaffold-based splits and applicability-domain analysis, and use regularization or dimension reduction to mitigate multicollinearity.

ARTICLE IN PRESS

Finally, the exponent-based framework suggests natural extensions to other entropy functionals, such as Rényi or Tsallis entropy, and to weighted, directed, or heterogeneous networks.

Conflict of interest

The authors declare that they have no conflict of interest in this paper.

Funding

No funding is available for this study.

Author contributions

Conceptualization: M.A., Y.Z.; Methodology: Y.Z., M.A.; Validation: M.A.; Investigation: M.A., Y.Z.; Resources: M.A.; Writing-original draft: M.A.; Writing-review and editing:Y.Z. Both authors have read and agreed to the published version of the manuscript.

Data Availability Statement

The data analyzed in this study are publicly available from the ChEMBL database (`https://www.ebi.ac.uk/chembl/`). All curated datasets, descriptor calculation scripts (classical and Prodeg entropy families, physicochemical properties, etc.), and modeling code used in the QSPR ablation studies have been provided as **Supplementary Materials** with this submission to support evaluation and reproducibility during peer review. Upon acceptance, we will make the complete reproducibility package publicly available via a Zenodo archived release (with a DOI) and update the manuscript with the corresponding permanent links.

ARTICLE IN PRESS

References

[1] Shannon, C. E. (1948). A mathematical theory of communication. *Bell System Technical Journal, 27*, 379-423, 623-656. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

[2] Rashevsky, N. (1955). Life, information theory, and topology. *Bulletin of Mathematical Biophysics, 17*, 229-235. https://doi.org/10.1007/BF02477860

[3] Körner, J. (1973). Coding of an information source having ambiguous alphabet and the entropy of graphs. In *Transactions of the Sixth Prague Conference on Information Theory, Statistical Decision Functions, Random Processes* (pp. 411-425).

[4] Simonyi, G. (1995). Graph entropy: A survey. In *Combinatorial optimization (New Brunswick, NJ, 1992-1993)* (DIMACS Series in Discrete Mathematics and Theoretical Computer Science, Vol. 20, pp. 399-441). American Mathematical Society. https://doi.org/10.1090/dimacs/020/08

ARTICLE IN PRESS

[5] Dehmer, M., & Mowshowitz, A. (2011). A history of graph entropy measures. *Information Sciences, 181*(1), 57-78. https://doi.org/10.1016/j.ins.2010.08.041

[6] Dehmer, M., Emmert-Streib, F., Chen, Z., Li, X., & Shi, Y. (Eds.). (2016). *Mathematical foundations and applications of graph entropy* (Quantitative and Network Biology, Vol. 6). Wiley-VCH.

[7] Li, X., & Wei, M. (2016). Graph entropy: Recent results and perspectives. In M. Dehmer, F. Emmert-Streib, Z. Chen, X. Li, & Y. Shi (Eds.), *Mathematical foundations and applications of graph entropy* (pp. 133-182). Wiley-VCH.

[8] Cao, S., Dehmer, M., & Shi, Y. (2014). Extremality of degree-based graph entropies. *Information Sciences, 278*, 22-33. https://doi.org/10.1016/j.ins.2014.03.133

[9] Ilić, A. (2016). On the extremal values of general degree-based graph entropies. *Information Sciences, 370-371*, 424-427.

[10] Ghalavand, A., Eliasi, M., & Ashrafi, A. R. (2019). First degree-based entropy of graphs. *Applied Mathematics and Computation, 59*, 37-46.

[11] Dong, Y., Qiao, S., Chen, B., Wan, P., & Zhang, S. (2021). Maximum values of degree-based entropies of bipartite graphs. *Applied Mathematics and Computation, 401*, 126094.

[12] Cambie, S., Dong, Y., & Mazzamurro, M. (2024). Extremal values of degree-based entropies of bipartite graphs. *Information Sciences, 676*, 120737. https://doi.org/10.1016/j.ins.2024.120737

[13] Cambie, S., & Mazzamurro, M. (2023). Resolution of Yan's conjecture on entropy of graphs. *MATCH Communications in Mathematical and in Computer Chemistry, 89*(2), 389-404. https://doi.org/10.46793/match.89-2.389C

[14] Bollobás, B., & Nikiforov, V. (2004). Degree powers in graphs with forbidden subgraphs. *Electronic Journal of Combinatorics, 11*, R42.

[15] Bollobás, B., & Nikiforov, V. (2012). Degree powers in graphs: The Erdős-Stone theorem. *Combinatorics, Probability and Computing, 21*, 89-105.

[16] Gutman, I., & Trinajstić, N. (1972). Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons. *Chemical Physics Letters, 17*, 535-538.

[17] Marshall, A. W., & Olkin, I. (1976). *Inequalities: Theory of majorization and its applications*. Academic Press.

[18] Karamata, J. (1932). Sur une inégalité relative aux fonctions convexes. *Publications de l'Institut Mathématique, 1*(1), 145-147.

[19] Dehmer, M. (2008). Information processing in complex networks: Graph entropy and information functionals. *Applied Mathematics and Computation, 201*(1-2), 82-94. https://doi.org/10.1016/j.amc.2007.12.010

ARTICLE IN PRESS

[20] Chen, Z., Dehmer, M., & Shi, Y. (2014). A note on distance-based graph entropies. *Entropy, 16*, 5416-5427.

[21] Ghorbani, M., Dehmer, M., & Zangi, S. (2018). Graph operations based on using distance-based graph entropies. *Applied Mathematics and Computation, 333*, 547-555. https://doi.org/10.1016/j.amc.2018.04.003

[22] Cao, S., Dehmer, M., & Kang, Z. (2017). Network entropies based on independent sets and matchings. *Applied Mathematics and Computation, 307*, 265-270.

[23] Wan, P., Chen, X., Tu, J., Dehmer, M., Zhang, S., & Emmert-Streib, F. (2020). On graph entropy measures based on the number of independent sets and matchings. *Information Sciences, 516*, 491-504.

[24] Wan, P., Tu, J., Dehmer, M., Zhang, S., & Emmert-Streib, F. (2019). Graph entropy based on the number of spanning forests of c-cyclic graphs. *Applied Mathematics and Computation, 363*, 124616.

[25] Alameri, A. , Alsharafi, M., Ali, E. & Gumaan, M. S. (2021). A note on Topological indices and coindices of disjunction and symmetric difference of graphs. Discrete Applied Mathematics, 304, 230-235. https://doi.org/10.1016/j.dam.2021.07.037

[26] Alsharafi, M., & Alameri, A. (2021). The F-index and coindex of V-Phenylenic Nanotubes and Nanotorus and their molecular complement graphs. *Nanosystems: Physics, Chemistry, Mathematics*, *12*(3), 263-270. https://doi.org/10.17586/2220-8054-2021-12-3-263-270.

[27] Alsharafi, M., Zeren, Y., & Alameri, A. (2021). The Second Hyper-Zagreb Index of Complement Graphs and Its Applications of Some Nano Structures. Asian Journal of Probability and Statistics 15 (4):54-75. https://doi.org/10.9734/ajpas/2021/v15i430364.

[28] Alsharafi, M., Akar, M., Zeren, Y., & Alameri, A. (2024). Degree-Based Topological Descriptors of Hexaphenylbenzene Molecule Graphs. Polycyclic Aromatic Compounds, 44(2), 1238–1257. https://doi.org/10.1080/10406638.2023.2190133

[29] Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American Chemical Society, 69*(1), 17-20. https://doi.org/10.1021/ja01193a005

[30] Randić, M. (1975). Characterization of molecular branching. *Journal of the American Chemical Society, 97*, 6609-6615.

[31] Bonchev, D., & Rouvray, D. H. (Eds.). (2005). *Complexity in chemistry, biology and ecology*. Springer.

[32] Manzoor, S., Siddiqui, M. K., & Ahmad, S. (2020). On entropy measures of molecular graphs using topological indices. *Arabian Journal of Chemistry, 13*, 6285-6298.

[33] Manzoor, S., Siddiqui, M. K., & Ahmad, S. (2022). On entropy measures of polycyclic hydroxy-chloroquine used for novel coronavirus (COVID-19) treatment. *Polycyclic Aromatic Compounds, 42*, 2947-2969.

ARTICLE IN PRESS

[34] Wang, X. L., Siddiqui, M. K., Kirmani, S. A. K., Manzoor, S., Ahmad, S., & Dhlamini, M. (2021). On topological analysis of entropy measures for silicon carbides networks. *Complexity, 2021*, 1-26.

[35] Huang, R., Siddiqui, M. K., Manzoor, S., Khalid, S., & Almotairi, S. (2022). On physical analysis of topological indices via curve fitting for the natural polymer of cellulose network. *The European Physical Journal Plus, 137*, 1-17.

[36] Kumar, V., & Das, S. (2023). On Nirmala indices-based entropy measures of silicon carbide network. *Iranian Journal of Mathematical Chemistry, 14*, 271-288.

[37] Ishfaq, F., Nadeem, M. F., & El-Bahy, Z. M. (2023). On topological indices and entropies of diamond structure. *International Journal of Quantum Chemistry, 123*, e27207.

[38] Furtula, B.; Gutman, I. A Forgotten Topological Index. *J. Math. Chem.* **2015**, *53*(4), 1184-1190.

[39] Alameri, A. , Al-Naggar, N., Al-Rumaima, M., Alsharafi M., Y-index of some graph operations,' *International Journal of Applied Engineering Research.*, 15 (2), 2020, pp. 173-179.

[40] Kulli, V. R. Nirmala index, *Int. J. Math. Trends Technol.* **67** (2021) 8–12.

[41] Gutman, I. Geometric approach to degree based topological indices, MATCH Commun. Math. Comput. Chem. 86(1) (2021) 11-16.

[42] Kulli, V. Dharwad indices. International Journal of Engineering Sciences & Research Technology., 10(4), 2021, doi: 10. 17-21. 10.29121/ijesrt.v10.i4.2021.2.

[43] X. Li, J. Zheng, *A unified approach to the extremal trees for different indices*, MATCH Commun. Math. Comput. Chem., **54**, 195-208. 2005.

[44] Barrière, L., Dalfó, C., Fiol, M. A., & Mitjana, M. (2009). The generalized hierarchical product of graphs. *Discrete Mathematics*, *309*, 3871–3881.

[45] Eliasi, M., & Iranmanesh, A. (2013). Hosoya polynomial of hierarchical product of graphs. *MATCH Communications in Mathematical and in Computer Chemistry*, *69*, 111–119.

[46] Advanced Chemistry Development, Inc. (ACD/Labs). ACD/ChemSketch, version 2024. Toronto, ON, Canada. Available at: `https://www.acdlabs.com/products/chemsketch/`. Accessed 2025-12-25.

[47] G. Landrum and RDKit contributors. RDKit: Open-source cheminformatics. Available at: `https://www.rdkit.org/`. Accessed 2025-12-25.

[48] Hagberg, A. A., Schult, D. A., Swart. P. J., Exploring network structure, dynamics, and function using NetworkX. In *Proceedings of the 7th Python in Science Conference (SciPy 2008)*, pages 11-15, 2008.

ARTICLE IN PRESS

ARTICLE IN PRESS

[49] Meurer, A., Smith, C. P., Paprocki, M. *et al.* SymPy: symbolic computing in Python. *PeerJ Computer Science*, 3:e103, 2017. doi:10.7717/peerj-cs.103.

[50] Pedregosa, F., Varoquaux, G., Gramfort, A., *et al.* Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825-2830, 2011.

[51] PySMILES developers. PySMILES: SMILES reader for NetworkX (software). Available at: `https://github.com/pckroon/pysmiles`. Accessed 2025-12-25.

[52] Hoerl, A. E., Kennard, R. W. Ridge regression: Biased estimation for nonorthogonal problems. *Technometrics*, 12(1):55-67, 1970.

[53] Stone. M. Cross-validatory choice and assessment of statistical predictions. *Journal of the Royal Statistical Society, Series B*, 36(2):111-147, 1974.

[54] B. Zdrazil, E. Felix, & et. al. (2024). The ChEMBL database in 2023: A drug discovery platform spanning multiple bioactivity data types and time periods. *Nucleic Acids Research*, *52*(D1), D1180-D1192. `https://doi.org/10.1093/nar/gkad1004`

[55] Y. Benjamini & Y. Hochberg. (1995). Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society. Series B (Methodological)*, *57*(1), 289-300. `https://doi.org/10.1111/j.2517-6161.1995.tb02031.x`

[56] J. B. Liu, L. Guan, J. Cao & L. Chen. (2025). Coherence Analysis for a Class of Polygon Networks With the Noise Disturbance. *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, *55*(7), 4718-4727. `https://doi.org/10.1109/TSMC.2025.3559326`

[57] J. B. Liu, C. Wang, S. Wang & et al. (2019). Zagreb Indices and Multiplicative Zagreb Indices of Eulerian Graphs. *Bulletin of the Malaysian Mathematical Sciences Society*, *42*, 67–78. `https://doi.org/10.1007/s40840-017-0463-2`

[58] R. R. Huang, S. Azam, A. Aslam & et al. (2025). Topological indices and QSPR modeling of gonalgia-associated drug molecules via M-polynomials. *The European Physical Journal E*, *48*, 63. `https://doi.org/10.1140/epje/s10189-025-00529-9`

[59] X. Zhang, S. Yousaf, A. Naeem, F. M. Tawfiq & A. Aslam. (2024). Analyzing topological descriptors of guar gum and its derivatives for predicting physical properties in carbohydrates. *Chemometrics and Intelligent Laboratory Systems*, *253*, 105203. `https://doi.org/10.1016/j.chemolab.2024.105203`

[60] Liu, J.-B., Wang, K., & Zhai, X. (2026). Statistical analysis and topological property of a class of fractal networks. *Fractals*, 2640001. https://doi.org/10.1142/S0218348X26400013

[61] Liu, J.-B., Yu, M.-Y., Cai, G.-J., & Cao, J. (2026). The leader–follower coherence and spectral properties of weighted recursive networks under noise disturbance. *Chaos, Solitons & Fractals*, *208*, 118060. https://doi.org/10.1016/j.chaos.2026.118060

ARTICLE IN PRESS