Scientific Reports

https://doi.org/10.1038/s41598-026-56190-5

Article in Press

Degree-based Inverse Prodeg indices linking graph theory and cheminformatics with sharp results and QSPR modeling

Received: 19 February 2026

Accepted: 29 May 2026

Published online: 12 June 2026

Cite this article as: Alsharafi M. & Zeren Y. Degree-based Inverse Prodeg indices linking graph theory and cheminformatics with sharp results and QSPR modeling. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-56190-5

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

Degree-Based Inverse Prodeg Indices Linking Graph Theory and Cheminformatics with Sharp Results and QSPR Modeling

Mohammed Alsharafi[a,b,*], **Yusuf Zeren**[b],

[a] *Department of Mathematics, Faculty of Science, Sana'a University, Sana'a, Yemen.* [b] *Department of Mathematics, Faculty of Arts and Science, Yildiz Technical University, 34220 Esenler, Istanbul, Turkey.*

alsharafi205010@gmail.com, yzeren@yildiz.edu.tr

(Received May 29, 2026)

Abstract

Degree-based topological indices are widely used in mathematical chemistry because they provide simple numerical descriptions of molecular graphs and can support structure–property analysis. In this work, we introduce the Inverse Prodeg index $IP(G)$ and its coindex $\overline{IP}(G)$ as new degree-derived graph invariants. Unlike product-, sum-, and mixed-degree descriptors such as the Randić, sum-connectivity, harmonic, atom-bond connectivity, geometric-arithmetic, Sombor, Nirmala, inverse Nirmala, and misbalance prodeg indices, $IP(G)$ reduces to the vertex-wise concave sum $\sum_{u\in V(G)} \sqrt{\delta_G(u)}$, whereas $\overline{IP}(G)$ transfers the same inverse square-root degree weighting to nonedges using the original graph degrees. We establish their main mathematical properties, including bounds involving graph order, size, and degree extrema, equality cases, Nordhaus–Gaddum-type inequalities, exact expressions for standard graph families, and estimates under several graph operations. These results show that the proposed descriptors are analytically tractable and computationally efficient, with linear-time computability in the number of edges. To examine their chemical relevance, we evaluate a small Prodeg-based descriptor family on a dataset of 90 aromatic-carboxylate compounds using training and external test sets generated by the Kennard-Stone algorithm. The models indicate that these descriptors capture useful structure-property information, especially for size- and thermodynamics-related endpoints. Linear regression and partial least-squares regression achieved the strongest average external-test performance among the considered Prodeg-only models, with mean external-test $R^2 = 0.832$ across the nine core endpoints, while Ridge regression was close with mean external-test $R^2 = 0.829$. Nonlinear methods did not improve the average prediction accuracy. Additional validation through bootstrap analysis, Y-randomization, residual diagnostics, and applicability-domain assessment supports a non-spurious but dataset-dependent predictive signal. Overall, the Inverse Prodeg index and its coindex provide mathematically well-founded and practically useful graph descriptors, although broader validation and combination with chemically richer descriptors are needed before claiming general predictive superiority.

Keywords: Inverse Prodeg index; coindex; degree-based topological descriptors; graph operations; QSPR; external validation.

ARTICLE IN PRESS

1 Introduction

Degree-based topological indices are foundational in chemical graph theory and network science, serving as powerful quantitative descriptors for molecular structures and complex networks [1–4]. The classical Zagreb indices, introduced by Gutman and Trinajstić [1] and further developed in subsequent work [2], remain central to this field owing to their simplicity and strong structure-property performance. These and many subsequent variants have been employed extensively in the prediction of physicochemical properties (QSPR/QSAR) [4,5], in the analysis of biological and social networks [6], and in the design of technological systems. A recurring motivation for new indices is increased flexibility, interpretability, or sensitivity to salient structural features (e.g., branching or the prevalence of low-degree vertices). The general Zagreb frameworks provide one such direction, introducing tunable parameters to better capture fine structural nuances [4,7].

Despite their success, many classical edge-weighted descriptors may respond similarly on graphs with distinct degree distributions, or they may overweight a small set of high-degree vertices when negative exponents are absent. In QSPR settings, such behavior can reduce interpretability when one seeks a descriptor that (i) preserves analytic tractability, (ii) remains sensitive to branching via the degree sequence, and (iii) yields clean extremal characterizations. These considerations motivate indices built from concave transforms of degrees, where majorization tools (e.g. Karamata-type arguments) become directly applicable.

Recent advances have focused on sharp inequalities, extremal characterizations, and Nordhaus–Gaddum-type relations linking a graph to its complement [5,7,8]. These results deepen the connection between a graph's degree sequence and global topology, with implications for both theory and applications. In particular, complement-based descriptors and coindices have received sustained attention in both theoretical developments and nanostructure applications [9–11].

Unless stated otherwise, all graphs in this paper are finite, simple (no loops or multiple edges), and undirected; connectedness is assumed whenever required by the statement of a result. Let $G = (V, E)$ have order $\mathfrak{n} = |V|$ and size

*Corresponding author

ARTICLE IN PRESS

$\mathfrak{m} = |E|$. For $u \in V$, we write $\delta_G(u)$ for its degree (we reserve $\delta(G)$ and $\Delta(G)$ for the minimum and maximum degree, respectively). An r-regular graph satisfies $\delta_G(u) = r$ for all $u \in V$. All formulas involving reciprocal degree terms, in particular the coindex $\overline{IP}(G)$, are stated for graphs with no isolated vertices unless the statement explicitly says otherwise; equivalently, $\delta(G) \geqslant 1$. This convention avoids undefined factors of the form $1/\sqrt{\delta_G(u)}$. The complement $\overline{G}$ has edge set consisting of all nonedges of G, and

$$\delta_{\overline{G}}(u) = \mathfrak{n} - 1 - \delta_G(u), \qquad \mathfrak{m} + |E(\overline{G})| = \binom{\mathfrak{n}}{2} = \frac{\mathfrak{n}(\mathfrak{n}-1)}{2}.$$

Degree-based topological indices are among the most widely used molecular descriptors in chemical graph theory and QSPR/QSAR modeling [12]. Classic representatives are the Zagreb indices, introduced to capture branching and substituent effects via degree sums and products over edges. The first Zagreb index is

$$M_1(G) = \sum_{uv \in E(G)} \big(\delta_G(u) + \delta_G(v)\big),$$

and the second Zagreb index is

$$M_2(G) = \sum_{uv \in E(G)} \delta_G(u)\, \delta_G(v),$$

both originally formulated in the early 1970s [1]. (Note that $M_1(G) = \sum_{v \in V(G)} \delta_G(v)^2$ and the edge form above are equivalent.)

A closely related family is the Randić connectivity index. The original Randić index (also called the product-connectivity index) is

$$R_{-1/2}(G) = \sum_{uv \in E(G)} \big(\delta_G(u)\, \delta_G(v)\big)^{-1/2},$$

designed to quantify molecular branching and found to correlate with several physicochemical properties [13]. This was generalized to the *general Randić* index

$$R^{\alpha}(G) = \sum_{uv \in E(G)} \big(\delta_G(u)\, \delta_G(v)\big)^{\alpha},$$

for a real exponent α [14]. In parallel, Zhou and Trinajstić introduced the *sum-connectivity* family

$$\chi^{\alpha}(G) = \sum_{uv \in E(G)} \big(\delta_G(u) + \delta_G(v)\big)^{\alpha},$$

whose special case $\alpha = -\frac{1}{2}$ emphasizes degree sums and often performs well in alkane series [15]. A limiting member is the *harmonic* index

$$H(G) = \sum_{uv \in E(G)} \frac{2}{\delta_G(u) + \delta_G(v)},$$

which enjoys elegant extremal properties and analytic tractability [16]. Collectively, these indices weight edges according to the degrees of their endpoints, with negative exponents down-weighting high-degree pairs and sum-based forms comparatively favoring edges incident to low-degree vertices.

Beyond these classics, several degree-ratio and degree-mixture descriptors have proved influential. The *atom-bond connectivity* index

$$ABC(G) = \sum_{uv \in E(G)} \sqrt{\frac{\delta_G(u) + \delta_G(v) - 2}{\delta_G(u)\, \delta_G(v)}}$$

was proposed to model heats of formation and strain energies [17]. The *geometric-arithmetic* index

$$GA(G) = \sum_{uv \in E(G)} \frac{2\sqrt{\delta_G(u)\, \delta_G(v)}}{\delta_G(u) + \delta_G(v)}$$

balances product and sum effects and has shown robust QSPR performance [18]. Local irregularity is captured by the *symmetric-division degree* index

$$SDD(G) = \sum_{uv \in E(G)} \frac{\max\{\delta_G(u), \delta_G(v)\}}{\min\{\delta_G(u), \delta_G(v)\}},$$

ARTICLE IN PRESS

which measures symmetric degree quotients [19]. More recently, the *Sombor* index

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

has attracted attention for its simple geometric form and strong empirical correlations [20]. We also note the *inverse sum indeg* index

$$\mathrm{ISI}(G) = \sum_{uv \in E(G)} \frac{\delta_G(u)\,\delta_G(v)}{\delta_G(u) + \delta_G(v)},$$

sometimes viewed as a Redefined second Zagreb index, with results on extremal values and spectral connections [21]. These descriptors enrich the palette of degree-based indices by mixing degrees through addition, multiplication, or more elaborate transforms to target specific structural features (branching, heterogeneity) relevant for QSPR.

Among more recent proposals, the *Nirmala* index by Kulli in [22] is defined by

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

which can be viewed as a sum-connectivity type descriptor with exponent $\frac{1}{2}$. The *misbalance prodeg* index (also called MPI) [22] is

$$MPI(G) = \sum_{uv \in E(G)} \left(\sqrt{\delta_G(u)} + \sqrt{\delta_G(v)}\right) = \sum_{u \in V(G)} \delta_G(u)\sqrt{\delta_G(u)} = \sum_{u \in V(G)} \delta_G(u)^{3/2},$$

which emphasizes higher degrees through the convex transform $x^{3/2}$ and thus responds differently to branching than harmonic- or Randić-type measures. Later in 2021, Kulli and coauthors introduced the *inverse Nirmala* indices [23],

$$IN_1(G) = \sum_{uv \in E(G)} \left(\frac{1}{\delta_G(u)} + \frac{1}{\delta_G(v)}\right)^{\frac{1}{2}}, \qquad IN_2(G) = \sum_{uv \in E(G)} \left(\frac{1}{\delta_G(u)} + \frac{1}{\delta_G(v)}\right)^{-\frac{1}{2}},$$

which complement $N(G)$ by coupling inverse degrees under square-root and its reciprocal, respectively.

Within this landscape, the *Inverse Prodeg* index introduced in this work offers a new perspective by combining inverse-degree weighting with a concave (square-root) transformation of degrees. We define

$$IP(G) = \sum_{uv \in E(G)} \left(\frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}}\right) = \sum_{u \in V(G)} \sqrt{\delta_G(u)}. \tag{1}$$

In particular, since $IP(G)$ depends only on the degree multiset, it is invariant under graph isomorphisms.

The equality in (1) follows because each u contributes $\delta_G(u) \cdot \left(1/\sqrt{\delta_G(u)}\right) = \sqrt{\delta_G(u)}$ across its incident edges. Unlike edge-coupled weights that depend symmetrically on both endpoints (e.g., ABC, GA, Randić, N, IN_1, IN_2), $IP(G)$ ultimately aggregates a vertex-wise concave transform. Consequently, $IP(G)$ is highly sensitive to degree-sequence majorization: by Karamata's inequality, for fixed $\sum_{u \in V(G)} \delta_G(u) = 2\mathfrak{m}$, the sum $\sum_{u \in V(G)} \sqrt{\delta_G(u)}$ is maximized by more *balanced* degree distributions and minimized by more *uneven* ones. This concavity yields clean extremal theory via majorization and Cauchy–Schwarz arguments.

The novelty of the present descriptors (Table 1) is therefore not merely the introduction of another degree formula, but the combination of four features: (i) an inverse square-root edge contribution that collapses exactly to a vertex-wise concave degree sum, (ii) direct sensitivity to degree-sequence balance and branching, (iii) a complementary nonedge version that keeps the degree information of the original molecular graph, and (iv) closed-form extremal behavior under standard graph families and graph operations. These features distinguish the proposed descriptors from both product-degree and sum-degree families and also make their chemical interpretation transparent.

On the considered case studies, $IP(G)$ performs competitively against several classical indices, but the empirical conclusions are necessarily limited by dataset size and should be interpreted as preliminary evidence motivating broader validation on larger benchmark sets.

ARTICLE IN PRESS

Table 1. Novelty of the proposed Inverse Prodeg descriptors relative to representative degree-based molecular descriptors.

Descriptor family	Main degree mechanism	Distinction from $IP(G)$ and $\overline{IP}(G)$
Randić/product-connectivity	Product of endpoint degrees, usually $(\delta_u\delta_v)^\alpha$	Edge-coupled product weighting; does not reduce to the concave vertex sum $\sum_u \sqrt{\delta_u}$.
Sum-connectivity/harmonic	Sum of endpoint degrees, e.g. $(\delta_u + \delta_v)^\alpha$ or $2/(\delta_u + \delta_v)$	Sum-based edge weighting; sensitive to endpoint pairs rather than to degree-sequence majorization through a vertex-wise square-root transform.
ABC, GA, SDD, Sombor	Mixed degree-ratio or geometric forms	Capture endpoint heterogeneity or Euclidean degree magnitude; their extremal theory generally uses different inequalities and does not produce the same complement coindex.
Nirmala/inverse Nirmala	Square-root transforms of degree sums or inverse-degree sums	Remain edge-pair descriptors; the proposed IP uses inverse square-root endpoint terms that collapse exactly to a vertex-level concave degree aggregate.
Misbalance prodeg (MPI)	Convex vertex contribution $\sum_u \delta_u^{3/2}$	Emphasizes high-degree/branched environments, whereas IP uses $\sum_u \delta_u^{1/2}$, which gives relatively stronger weight to low-degree and terminal environments.
Proposed $IP, \overline{IP}$	Concave vertex degree sum and nonedge coindex using original degrees	Analytically tractable, low-degree sensitive, compatible with Nordhaus–Gaddum/coindex analysis, and directly interpretable for hydrogen-suppressed molecular graphs.

The corresponding coindex is defined as a nonedge-based sum that uses degrees taken in G:

$$\overline{IP}(G) = \sum_{uv\in E(\overline{G})} \left(\frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}} \right). \tag{2}$$

Remark. The Inverse Prodeg coindex $\overline{IP}(G)$ should not be confused with the Inverse Prodeg index of the complement $IP(\overline{G})$. The former is defined by (2) and uses degrees in G, whereas

$$IP(\overline{G}) = \sum_{uv\in E(\overline{G})} \left(\frac{1}{\sqrt{\delta_{\overline{G}}(u)}} + \frac{1}{\sqrt{\delta_{\overline{G}}(v)}} \right). \tag{3}$$

In general, $\overline{IP}(G) \neq IP(\overline{G})$; they coincide only for special graph families (e.g. some regular or highly symmetric graphs).

ARTICLE IN PRESS

Table 2. Notation used for the Inverse Prodeg index, its coindex, and the complement graph.

Notation	Meaning
$IP(G)$	Inverse Prodeg index of G, summed over $E(G)$.
$\overline{IP}(G)$	Inverse Prodeg coindex, summed over $E(\overline{G})$ but using degrees in G.
$IP(\overline{G})$	Inverse Prodeg index of the complement graph, using degrees in $\overline{G}$.
$\delta_G(u)$, $\delta(G)$, $\Delta(G)$	Degree of u, minimum degree, and maximum degree in G.
$\mathfrak{n}, \mathfrak{m}, \overline{\mathfrak{m}}$	Order, size, and complement size of G.

More generally, for $\alpha \geqslant 0$ we consider

$$IP^\alpha(G) = \sum_{uv\in E(G)} \left((\delta_G(u))^{-\alpha/2} + (\delta_G(v))^{-\alpha/2} \right) = \sum_{uv\in E(G)} \left(\frac{1}{\sqrt{(\delta_G(u))^\alpha}} + \frac{1}{\sqrt{(\delta_G(v))^\alpha}} \right),$$

with the analogous coindex $\overline{IP}^\alpha(G)$ defined by summing over $uv \in E(\overline{G})$. Special cases include

$$IP^0(G) = 2\mathfrak{m},$$

$$IP^2(G) = \sum_{uv\in E(G)} \left(\tfrac{1}{\delta_G(u)} + \tfrac{1}{\delta_G(v)} \right) = \sum_{uv\in E(G)} \left(\tfrac{\delta_G(u)+\delta_G(v)}{\delta_G(u)\delta_G(v)} \right) = ReZ_1(G),$$

ARTICLE IN PRESS

$$IP^4(G) = \sum_{uv \in E(G)} \left(\frac{1}{\delta_G(u)^2} + \frac{1}{\delta_G(v)^2} \right),$$

recovering, respectively, a trivial edge-counting limit, the edge-based inverse-degree sum, and the inverse-square-degree sum used in the literature.

Graph operations-such as disjoint union, join, complement, Cartesian and lexicographic (composition) products, strong and tensor products, corona, rooted product, and hierarchical/generalized hierarchical products-are a standard device in chemical graph theory for building complex networks from simpler components and for transferring structural information between families of graphs. A substantial literature develops closed forms, recurrences, and bounds for classical indices under these operations. Foundational contributions by Khalifeh, Yousefi-Azari, and Ashrafi derived exact expressions for the hyper-Wiener index and for the first and second Zagreb indices across several operations [24, 25], while Ashrafi, Došlić, and Hamzeh introduced and analyzed Zagreb *coindices* in the same setting [26]. Multiplicative Zagreb indices under graph operations were investigated by Das and coauthors [27]. Azari and Iranmanesh studied distance-based descriptors-most notably the eccentric–distance sum (EDS)-for a variety of operations [28], obtained Zagreb index formulas for rooted products [29], established sharp lower bounds for the Narumi–Katayama index under operations [30], and proved inequalities for the multiplicative sum Zagreb index [31]. Measures of irregularity and eccentric connectivity on composite graphs and operation-generated families have been analyzed in [32–35], and higher-order Zagreb (co)indices, including the third and hyper–Zagreb coindices, have been derived for standard operations [36]. Product constructions tailored to hierarchical modeling are covered by the generalized hierarchical product [37] and its Hosoya polynomial [38]. More recently, complement–based operations and coindex-type constructions have been explored for hyper-Zagreb and related descriptors [10,39,40], while disjunction and symmetric difference operations have also been investigated from the viewpoint of topological indices and coindices [41].

In line with this body of work, we analyze $IP(G)$ under the join, Cartesian product, strong product, tensor product, disjunction, and symmetric difference, providing exact formulas in cases where they are available (e.g. tensor products) and sharp bounds in the remaining cases.

Quantitative structure-property relationship (QSPR) modelling that leverages graph-theoretic descriptors has become a staple for predicting physicochemical, pharmacological, and ADMET properties (absorption, distribution, metabolism, excretion, and toxicity) of drug molecules. A substantial body of recent work has focused on degree-based and neighborhood degree-based indices to encode branching and local environments in molecular graphs. For antituberculosis agents, Adnan et al. demonstrated that degree-based indices yield strong predictive performance in QSPR analyses [42]. Balasubramaniyan and Chidambaram employed neighborhood degree-based descriptors for asthma drugs, highlighting the added value of local degree information [43]. Gnanaraj et al. extended these approaches to nonsteroidal anti-inflammatory drugs (NSAIDs), uncovering significant structure-property relationships [44]. For emergent indications, Kalaimathi and Balamurugan modeled physicochemical and ADMET parameters of candidate monkeypox therapeutics via molecular graph representations [45], while Kansal et al. introduced temperature-based indices to model COVID-19 drugs, illustrating methodological adaptability to rapidly evolving therapeutic contexts [46]. Separately, Hayat et al. developed novel temperature-based indices to model properties of polycyclic aromatic hydrocarbons and silicon carbide nanotubes [47,48]. Most recently, Manonmani and Selvarani combined multilinear regression with topological indices to analyze antiviral medications, improving the precision of property predictions [49]. Recent work has also tested structure sensitivity and chemical applicability of novel degree-based indices and applied degree-based regression models to anticancer or blood-cancer drug datasets [67, 68]. Earlier, Syed et al. reported successful QSPR modelling for breast-cancer drugs using degree-centric descriptors [50]. In addition, degree-sensitive indices such as the hyper-Zagreb and Sombor families have been shown to correlate well with experimental properties in QSPR settings, including applications to trees, monocarboxylic acids, and related molecular classes [51, 52].

Recent studies further illustrate the current movement toward descriptor–QSPR integration: Arockiaraj et al. used integrated face-degree topological indices for QSPR modeling of aluminum and magnesium hydroxide networks [53]; Kara et al. combined Sombor topological indices with entropy measures for QSPR modeling of anticancer drugs [54]; Henry et al. integrated QSAR/QSPR, clustering, and docking for oxadiazole anticancer candidates [55]; and topological characterization has also been applied to kekulene tessellations and related nanostructures [56].

Recent QSPR/QSAR studies increasingly combine molecular descriptors with nonlinear and interpretable machine-learning models, including XGBoost, random forests, support vector machines, and neural-network architectures. For example, recent work has used FNN, XGBoost, and RF models with SHAP interpretation for octanol–water partition prediction [57], graph-theoretic descriptors with regression and feedforward ANN models for drug-property prediction

ARTICLE IN PRESS

[58], and XGBoost/ensemble models with SHAP analysis for bioactivity or toxicity-related QSAR tasks [59, 60]. These developments show that modern QSPR research is moving toward nonlinear prediction together with explainable feature-attribution tools. In the present study, the Prodeg descriptors are therefore evaluated mainly as compact, interpretable graph-theoretic variables; their integration with larger descriptor pools, XGBoost-type models, neural networks, and SHAP-based interpretation is a natural direction for future work.

From a chemical-informatics viewpoint, degree-based descriptors quantify local connectivity and branching in molecular graphs, where the vertex degree reflects the valency/neighbor environment of an atom. The proposed Inverse Prodeg (IP) index incorporates inverse square-root degree weights, thereby assigning relatively larger contributions to low-degree (terminal) atoms and less-substituted environments. Such structural features are closely tied to functional group exposure, molecular branching, and local heterogeneity, which are frequently implicated in variations of physicochemical and ADMET-related properties. This evidence base motivates our introduction of the Inverse Prodeg index as a simple, interpretable, and theoretically tractable degree-based descriptor; in the sequel we place it within the broader index landscape and quantify its predictive value.

By (1), $IP(G)$ aggregates a concave transform of the degree sequence, which makes it particularly sensitive to branching (degree-sequence majorization) while remaining analytically tractable (via Karamata or Cauchy-Schwarz arguments). The coindex $\overline{IP}(G)$ transfers the same local weighting to the complement, enabling Nordhaus-Gaddum-type analyses. Next, we develop sharp bounds, extremal results, and NG-type inequalities for IP and $\overline{IP}$, derive closed forms for classical graph families, and present externally validated QSPR experiments on an aromatic-carboxylate compound dataset.

2 Specific Graph Classes

We now determine the values of $IP(G)$ and $\overline{IP}(G)$ for some classical graph families. For clarity, we recall the graph classes used in this section. $K_{\mathfrak{n}}$ is the complete graph, $S_{\mathfrak{n}} = K_{1,\mathfrak{n}-1}$ is the star, $P_{\mathfrak{n}}$ and $C_{\mathfrak{n}}$ denote the path and cycle, $K_{\mathfrak{m},\mathfrak{n}}$ is the complete bipartite graph, $K_{\mathfrak{n}_1,\ldots,\mathfrak{n}_k}$ is the complete multipartite graph, $W_{\mathfrak{n}}$ is obtained by joining one hub to all vertices of $C_{\mathfrak{n}-1}$, Q_d is the d-dimensional hypercube, $F_{\mathfrak{n}}$ is the friendship graph formed by $\mathfrak{n}$ triangles sharing one common vertex, and $L_{\mathfrak{n}} = P_2 \square P_{\mathfrak{n}}$ is the ladder graph. These are standard graph families; see, for example, [61, 62].
Recall that $\overline{IP}(G)$ is defined by summing over $E(\overline{G})$ but using degrees in G:

$$\overline{IP}(G) = \sum_{uv \in E(\overline{G})} \left(\frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}} \right).$$

Equivalently, since each vertex u is incident with $\mathfrak{n} - 1 - \delta_G(u)$ nonedges, we have

$$\overline{IP}(G) = \sum_{u \in V(G)} \frac{\mathfrak{n} - 1 - \delta_G(u)}{\sqrt{\delta_G(u)}}.$$

Proposition 1. *For the complete graph $K_{\mathfrak{n}}$, we have*

$$IP(K_{\mathfrak{n}}) = \mathfrak{n}\sqrt{\mathfrak{n}-1} \qquad \textit{and} \qquad \overline{IP}(K_{\mathfrak{n}}) = 0.$$

Proof. In $K_{\mathfrak{n}}$, every vertex has degree $\delta_G(u) = \mathfrak{n} - 1$, hence

$$IP(K_{\mathfrak{n}}) = \sum_{u \in V(K_{\mathfrak{n}})} \sqrt{\delta_G(u)} = \mathfrak{n}\sqrt{\mathfrak{n}-1}.$$

Moreover, $E(\overline{K_{\mathfrak{n}}}) = \varnothing$, so $\overline{IP}(K_{\mathfrak{n}}) = 0$. ■

Proposition 2. *For the star graph $S_{\mathfrak{n}}$, we have*

$$IP(S_{\mathfrak{n}}) = \sqrt{\mathfrak{n}-1} + \mathfrak{n} - 1 \qquad \textit{and} \qquad \overline{IP}(S_{\mathfrak{n}}) = (\mathfrak{n}-1)(\mathfrak{n}-2).$$

Proof. In $S_{\mathfrak{n}}$, one vertex has degree $\mathfrak{n} - 1$ and the remaining $\mathfrak{n} - 1$ vertices have degree 1. Thus,

$$IP(S_{\mathfrak{n}}) = \sqrt{\mathfrak{n}-1} + (\mathfrak{n}-1)\sqrt{1} = \sqrt{\mathfrak{n}-1} + \mathfrak{n} - 1.$$

ARTICLE IN PRESS

The nonedges are exactly the pairs of distinct leaves. There are $\binom{n-1}{2}$ such pairs, and each leaf has degree 1 in G; hence each nonedge contributes $1 + 1 = 2$. Therefore,

$$\overline{IP}(S_n) = \binom{n-1}{2} \cdot 2 = (n-1)(n-2).$$

■

Proposition 3. *For the cycle graph C_n, we have*

$$IP(C_n) = n\sqrt{2},$$

and for $n \geqslant 4$,

$$\overline{IP}(C_n) = \frac{n(n-3)}{\sqrt{2}}.$$

Proof. Each vertex of C_n has degree 2, hence

$$IP(C_n) = \sum_{u \in V(C_n)} \sqrt{2} = n\sqrt{2}.$$

For $n \geqslant 4$, the number of nonedges is $\binom{n}{2} - n = \frac{n(n-3)}{2}$. Each endpoint has degree 2 in G, so each nonedge contributes $2/\sqrt{2} = \sqrt{2}$. Thus,

$$\overline{IP}(C_n) = \left(\frac{n(n-3)}{2}\right)\sqrt{2} = \frac{n(n-3)}{\sqrt{2}}.$$

■

Proposition 4. *For the path graph P_n, we have*

$$IP(P_n) = 2 + \sqrt{2}\,(n-2).$$

Proof. The two end vertices have degree 1 and the $n-2$ internal vertices have degree 2. Hence,

$$IP(P_n) = 2\sqrt{1} + (n-2)\sqrt{2} = 2 + \sqrt{2}\,(n-2).$$

■

Proposition 5. *For the complete bipartite graph $K_{m,n}$, we have*

$$IP(K_{m,n}) = m\sqrt{n} + n\sqrt{m} = mn\left(\frac{1}{\sqrt{m}} + \frac{1}{\sqrt{n}}\right),$$

and

$$\overline{IP}(K_{m,n}) = \frac{m(m-1)}{\sqrt{n}} + \frac{n(n-1)}{\sqrt{m}}.$$

Proof. Vertices in the m-part have degree n, and vertices in the n-part have degree m. Hence,

$$IP(K_{m,n}) = m\sqrt{n} + n\sqrt{m}.$$

Nonedges occur only within the parts. A nonedge inside the m-part has endpoints of degree n in G, hence contributes $2/\sqrt{n}$, and there are $\binom{m}{2}$ such pairs. Similarly, a nonedge inside the n-part contributes $2/\sqrt{m}$ and there are $\binom{n}{2}$ such pairs. Therefore,

$$\overline{IP}(K_{m,n}) = \binom{m}{2}\frac{2}{\sqrt{n}} + \binom{n}{2}\frac{2}{\sqrt{m}} = \frac{m(m-1)}{\sqrt{n}} + \frac{n(n-1)}{\sqrt{m}}.$$

■

Proposition 6. *If G is an r-regular graph on n vertices, then*

$$IP(G) = n\sqrt{r} \qquad \text{and} \qquad \overline{IP}(G) = \frac{n(n-1-r)}{\sqrt{r}}.$$

ARTICLE IN PRESS

ARTICLE IN PRESS

Proof. Since $\delta_G(u) = r$ for all $u \in V(G)$,

$$IP(G) = \sum_{u \in V(G)} \sqrt{r} = \mathfrak{n}\sqrt{r}.$$

Also, each vertex is nonadjacent to $\mathfrak{n} - 1 - r$ vertices, so the identity yields

$$\overline{IP}(G) = \sum_{u \in V(G)} \frac{\mathfrak{n} - 1 - r}{\sqrt{r}} = \frac{\mathfrak{n}(\mathfrak{n} - 1 - r)}{\sqrt{r}}.$$

■

Proposition 7. *For the complete multipartite graph $K_{\mathfrak{n}_1,\mathfrak{n}_2,\ldots,\mathfrak{n}_k}$ with $\mathfrak{n} = \sum_{i=1}^{k} \mathfrak{n}_i$, we have*

$$IP(K_{\mathfrak{n}_1,\ldots,\mathfrak{n}_k}) = \sum_{i=1}^{k} \mathfrak{n}_i \sqrt{\mathfrak{n} - \mathfrak{n}_i}.$$

Moreover, equivalently,

$$IP(K_{\mathfrak{n}_1,\ldots,\mathfrak{n}_k}) = \sum_{1 \leqslant i < j \leqslant k} \mathfrak{n}_i \mathfrak{n}_j \left(\frac{1}{\sqrt{\mathfrak{n} - \mathfrak{n}_i}} + \frac{1}{\sqrt{\mathfrak{n} - \mathfrak{n}_j}} \right).$$

Proof. Each vertex in part i is adjacent to all vertices outside part i, hence has degree $\mathfrak{n} - \mathfrak{n}_i$. Summing $\sqrt{\delta_G(u)}$ over all vertices gives the first formula; regrouping by pairs of parts yields the equivalent edge form. ■

Proposition 8. *For the wheel graph $W_{\mathfrak{n}}$ ($\mathfrak{n} \geqslant 4$), we have*

$$IP(W_{\mathfrak{n}}) = \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{3},$$

and

$$\overline{IP}(W_{\mathfrak{n}}) = \frac{(\mathfrak{n} - 1)(\mathfrak{n} - 4)}{\sqrt{3}}$$

Proof. The wheel $W_{\mathfrak{n}}$ consists of a cycle $C_{\mathfrak{n}-1}$ together with a central vertex adjacent to all rim vertices. Thus the central vertex has degree $\mathfrak{n} - 1$ and each rim vertex has degree 3, giving

$$IP(W_{\mathfrak{n}}) = \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{3}.$$

Nonedges occur only among rim vertices that are not adjacent on $C_{\mathfrak{n}-1}$. Their number is $\binom{\mathfrak{n}-1}{2} - (\mathfrak{n} - 1) = \frac{(\mathfrak{n}-1)(\mathfrak{n}-4)}{2}$, and each such nonedge contributes $2/\sqrt{3}$. Hence,

$$\overline{IP}(W_{\mathfrak{n}}) = \left(\frac{(\mathfrak{n} - 1)(\mathfrak{n} - 4)}{2} \right) \frac{2}{\sqrt{3}} = \frac{(\mathfrak{n} - 1)(\mathfrak{n} - 4)}{\sqrt{3}}.$$

■

Proposition 9. *For the hypercube Q_d, we have*

$$IP(Q_d) = 2^d \sqrt{d}.$$

Proof. A d-dimensional hypercube has $\mathfrak{n} = 2^d$ vertices and is d-regular. Hence $IP(Q_d) = \mathfrak{n}\sqrt{d} = 2^d\sqrt{d}$. ■

Proposition 10. *For the friendship graph $F_{\mathfrak{n}}$, we have*

$$IP(F_{\mathfrak{n}}) = \sqrt{2\mathfrak{n}} + 2\mathfrak{n}\sqrt{2}.$$

Proof. The graph $F_{\mathfrak{n}}$ consists of $\mathfrak{n}$ triangles sharing a common vertex. Thus it has $2\mathfrak{n} + 1$ vertices; the central vertex has degree $2\mathfrak{n}$ and the remaining $2\mathfrak{n}$ vertices have degree 2. Therefore,

$$IP(F_{\mathfrak{n}}) = \sqrt{2\mathfrak{n}} + 2\mathfrak{n}\sqrt{2}.$$

■

ARTICLE IN PRESS

Proposition 11. *For the ladder graph $L_{\mathfrak{n}} = P_2 \square P_{\mathfrak{n}}$, we have*

$$IP(L_{\mathfrak{n}}) = 4\sqrt{2} + (2\mathfrak{n} - 4)\sqrt{3}.$$

Proof. The ladder graph has $2\mathfrak{n}$ vertices. It has four corner vertices of degree 2 and the remaining $2\mathfrak{n}-4$ vertices of degree 3. Summing $\sqrt{\delta_G(u)}$ over all vertices gives the stated expression. ■

Corollary. *Let T be a tree with $\mathfrak{n} \geqslant 2$ vertices. Then $IP(T)$ is minimized by the star $S_{\mathfrak{n}}$ and maximized by the path $P_{\mathfrak{n}}$, that is,*

$$IP(S_{\mathfrak{n}}) \leqslant IP(T) \leqslant IP(P_{\mathfrak{n}}).$$

Proof. Every tree has degree sum $2(\mathfrak{n}-1)$ and at least two leaves. Since $x \mapsto \sqrt{x}$ is concave, transferring one unit of degree from a larger degree to a smaller positive degree cannot decrease the sum of square roots. Hence the most balanced tree degree sequence, $(2, \ldots, 2, 1, 1)$, gives the maximum and corresponds to $P_{\mathfrak{n}}$, whereas the most concentrated tree degree sequence, $(\mathfrak{n}-1, 1, \ldots, 1)$, gives the minimum and corresponds to $S_{\mathfrak{n}}$. ■

Remark. Let G be a simple graph and let $u, v \in V(G)$. Consider adding the edge uv (if absent) to obtain $G^+ = G + uv$. Then

$$IP(G^+) - IP(G) = \big(\sqrt{\delta_G(u)+1} - \sqrt{\delta_G(u)}\big) + \big(\sqrt{\delta_G(v)+1} - \sqrt{\delta_G(v)}\big),$$

and hence the increment is bounded by

$$0 < IP(G^+) - IP(G) \leqslant \frac{1}{2\sqrt{\delta_G(u)}} + \frac{1}{2\sqrt{\delta_G(v)}},$$

whenever $\delta_G(u), \delta_G(v) \geqslant 1$, using concavity of $\sqrt{x}$. Similarly, if $uv \in E(G)$ and $G^- = G - uv$, then

$$\begin{aligned} IP(G) - IP(G^-) &= \big(\sqrt{\delta_G(u)} - \sqrt{\delta_G(u)-1}\big) + \big(\sqrt{\delta_G(v)} - \sqrt{\delta_G(v)-1}\big) \\ &\leqslant \frac{1}{2\sqrt{\delta_G(u)-1}} + \frac{1}{2\sqrt{\delta_G(v)-1}}, \end{aligned}$$

whenever $\delta_G(u), \delta_G(v) \geqslant 2$. Thus, the sensitivity of $IP(G)$ to a single edge flip decreases as local degrees increase.

ARTICLE IN PRESS

3 Bounds for Inverse Prodeg Index and Coindex in Terms of Order, Size, Degree Extremes, and Topological Indices

Before presenting the bounds, we summarize the proof strategy to improve reproducibility. First, the edge definition of $IP(G)$ is converted to the vertex identity $IP(G) = \sum_{u \in V(G)} \sqrt{\delta_G(u)}$. Second, degree-extreme bounds follow by applying $\sqrt{\delta(G)} \leqslant \sqrt{\delta_G(u)} \leqslant \sqrt{\Delta(G)}$ vertex-wise or by counting how many times each vertex degree appears in edge/nonedge sums. Third, order-size bounds use the handshaking relation $\sum_u \delta_G(u) = 2\mathfrak{m}$ together with Jensen, Cauchy–Schwarz, or Karamata-type concavity arguments. Finally, the coindex bounds follow from the identity $\overline{IP}(G) = \sum_{u \in V(G)} (\mathfrak{n} - 1 - \delta_G(u))/\sqrt{\delta_G(u)}$, which makes explicit the nonedge multiplicity of each vertex.

In this section, we derive bounds for the Inverse Prodeg index and its coindex in terms of the order $\mathfrak{n}$, size $\mathfrak{m}$, degree extrema $\delta(G), \Delta(G)$, and related degree-based indices.

Theorem 12. *Let G be a simple graph of order $\mathfrak{n} \geqslant 2$ with $\mathfrak{m} = |E(G)|$ and $\overline{\mathfrak{m}} = |E(\overline{G})|$. Then:*

(I)

$$\frac{2\mathfrak{m}}{\sqrt{\mathfrak{n}-1}} \leqslant IP(G) \leqslant 2\mathfrak{m},$$

(II)

$$\frac{2\overline{\mathfrak{m}}}{\sqrt{\mathfrak{n}-1}} \leqslant \overline{IP}(G) \leqslant 2\overline{\mathfrak{m}}.$$

Proof. (I) For each $uv \in E(G)$, we have $1 \leqslant \delta_G(u), \delta_G(v) \leqslant \mathfrak{n}-1$, hence $(\mathfrak{n}-1)^{-1/2} \leqslant \delta_G(u)^{-1/2} \leqslant 1$ and similarly for v. Therefore,

$$\frac{2}{\sqrt{\mathfrak{n}-1}} \leqslant \frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}} \leqslant 2.$$

ARTICLE IN PRESS

Summing over $E(G)$ yields $\frac{2\mathfrak{m}}{\sqrt{\mathfrak{n}-1}} \leqslant IP(G) \leqslant 2\mathfrak{m}$.

(II) The same estimate holds for each nonedge $uv \in E(\overline{G})$ because it also satisfies $1 \leqslant \delta_G(u), \delta_G(v) \leqslant \mathfrak{n}-1$ (note that degrees are still taken in G in the definition of $\overline{IP}(G)$). Summing over $E(\overline{G})$ gives $\frac{2\overline{\mathfrak{m}}}{\sqrt{\mathfrak{n}-1}} \leqslant \overline{IP}(G) \leqslant 2\overline{\mathfrak{m}}$. ■

Theorem 13. *Let $G=(V,E)$ be a simple graph with $\mathfrak{n}=|V| \geqslant 2$, $\mathfrak{m}=|E|$, and minimum degree $\delta(G)$. Then*

$$\mathfrak{n}\sqrt{\delta(G)} \;\leqslant\; IP(G) \;\leqslant\; \sqrt{2\mathfrak{n}\mathfrak{m}}.$$

Moreover, equality holds on the left if and only if G is regular of degree $\delta(G)$, and on the right if and only if G is regular.

Proof. Lower bound. Since $\delta_G(u) \geqslant \delta(G)$ for every $u \in V$ and $x \mapsto \sqrt{x}$ is increasing,

$$IP(G) = \sum_{u\in V} \sqrt{\delta_G(u)} \;\geqslant\; \sum_{u\in V} \sqrt{\delta(G)} \;=\; \mathfrak{n}\sqrt{\delta(G)}.$$

Equality holds iff $\delta_G(u)=\delta(G)$ for all u, i.e. G is regular.

Upper bound. By Cauchy-Schwarz,

$$\Big(\sum_{u\in V} \sqrt{\delta_G(u)}\Big)^2 \leqslant \Big(\sum_{u\in V} 1^2\Big)\Big(\sum_{u\in V} \delta_G(u)\Big) = \mathfrak{n}\cdot 2\mathfrak{m},$$

so $IP(G) \leqslant \sqrt{2\mathfrak{n}\mathfrak{m}}$. Equality holds iff $\sqrt{\delta_G(u)}$ is constant over V, i.e. G is regular. ■

Theorem 14. *Let $G=(V,E)$ be a simple graph with $\mathfrak{m}=|E|$ and minimum and maximum degrees $\delta(G)$, $\Delta(G)$, and assume $\delta(G) \geqslant 1$. Then*

$$\frac{2\mathfrak{m}}{\sqrt{\Delta(G)}} \;\leqslant\; IP(G) \;\leqslant\; \frac{2\mathfrak{m}}{\sqrt{\delta(G)}},$$

with equality in either inequality if and only if G is regular.

Proof. For every edge $uv \in E$,

$$\frac{2}{\sqrt{\Delta(G)}} \leqslant \frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}} \leqslant \frac{2}{\sqrt{\delta(G)}},$$

since $\delta(G) \leqslant \delta_G(\cdot) \leqslant \Delta(G)$ and $x \mapsto x^{-1/2}$ is decreasing. Summing over E and using

$$\sum_{uv\in E}\Big(\tfrac{1}{\sqrt{\delta_G(u)}}+\tfrac{1}{\sqrt{\delta_G(v)}}\Big) = \sum_{u\in V}\frac{\delta_G(u)}{\sqrt{\delta_G(u)}} = \sum_{u\in V}\sqrt{\delta_G(u)} = IP(G),$$

yields the bounds. Equality forces $\delta_G(u)$ constant on V, i.e. G regular. ■

Theorem 15. *Let $G=(V,E)$ be a simple graph on $\mathfrak{n} \geqslant 2$ vertices with $\overline{\mathfrak{m}}=|E(\overline{G})|$, and assume $\delta(G) \geqslant 1$. Then*

$$\frac{2\overline{\mathfrak{m}}}{\sqrt{\Delta(G)}} \;\leqslant\; \overline{IP}(G) \;\leqslant\; \frac{2\overline{\mathfrak{m}}}{\sqrt{\delta(G)}},$$

with equality in either inequality if and only if G is regular.

Proof. For every nonedge $uv \in E(\overline{G})$ we still have $\delta(G) \leqslant \delta_G(u), \delta_G(v) \leqslant \Delta(G)$, hence

$$\frac{2}{\sqrt{\Delta(G)}} \leqslant \frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}} \leqslant \frac{2}{\sqrt{\delta(G)}}.$$

Summing over $E(\overline{G})$ gives the claimed bounds for $\overline{IP}(G)$. Equality forces $\delta_G(u)$ constant, hence G is regular. ■

Theorem 16. *Let G be a simple graph on $\mathfrak{n} \geqslant 2$ vertices. Then*

$$\mathfrak{n}\sqrt{\mathfrak{n}-1} \;\leqslant\; IP(G)+\overline{IP}(G) \;\leqslant\; \mathfrak{n}(\mathfrak{n}-1).$$

Moreover, equality on the left holds iff $G \cong K_{\mathfrak{n}}$; equality on the right holds iff $\mathfrak{n}=2$ (i.e. $G \cong K_2$).

Proof. Lower bound. Since

$$IP(G)+\overline{IP}(G) = \sum_{u\in V}\sqrt{\delta_G(u)} + \sum_{u\in V}\frac{\mathfrak{n}-1-\delta_G(u)}{\sqrt{\delta_G(u)}} = \sum_{u\in V}\left(\sqrt{\delta_G(u)} + \frac{\mathfrak{n}-1-\delta_G(u)}{\sqrt{\delta_G(u)}}\right),$$

ARTICLE IN PRESS

define $g(x) = \sqrt{x} + \dfrac{\mathfrak{n}-1-x}{\sqrt{x}}$ on the domain $x \in [1, \mathfrak{n}-1]$. A direct computation gives

$$g'(x) = \frac{3x-(\mathfrak{n}-1)}{2x^{3/2}}, \qquad g''(x) = \frac{3(\mathfrak{n}-1-x)}{4x^{5/2}} \geqslant 0,$$

so g is convex on $[1, \mathfrak{n}-1]$ and attains its minimum at the critical point $x = (\mathfrak{n}-1)/3$ or at the endpoints. In particular, the simple estimate $g(x) \geqslant \sqrt{\mathfrak{n}-1}$ holds for all $x \in [1, \mathfrak{n}-1]$ (with equality only at $x = \mathfrak{n}-1$). Therefore,

$$IP(G) + \overline{IP}(G) \geqslant \sum_{u \in V} \sqrt{\mathfrak{n}-1} = \mathfrak{n}\sqrt{\mathfrak{n}-1}.$$

Equality forces $\delta_G(u) = \mathfrak{n}-1$ for all u, i.e. $G \cong K_{\mathfrak{n}}$.

Upper bound. For every integer $t \geqslant 1$, $\sqrt{t} \leqslant t$. Hence

$$IP(G) = \sum_{u \in V} \sqrt{\delta_G(u)} \leqslant \sum_{u \in V} \delta_G(u) = 2\mathfrak{m}, \qquad \overline{IP}(G) \leqslant 2\overline{\mathfrak{m}}.$$

Since $\mathfrak{m} + \overline{\mathfrak{m}} = \binom{\mathfrak{n}}{2}$, we get

$$IP(G) + \overline{IP}(G) \leqslant 2(\mathfrak{m} + \overline{\mathfrak{m}}) = \mathfrak{n}(\mathfrak{n}-1).$$

For equality we need $\sqrt{\delta_G(u)} = \delta_G(u)$ for all u, which forces $\delta_G(u) \in \{1\}$ (since G is simple and $\delta_G(u) \geqslant 1$ in connected cases), and simultaneously $\overline{\mathfrak{m}} = 0$, which is possible only when $\mathfrak{n} = 2$. ■

Theorem 17. *Let G be a simple graph on $\mathfrak{n} \geqslant 2$ vertices such that $\delta_G(u) \leqslant \left\lfloor \frac{\mathfrak{n}-1}{2} \right\rfloor$ for all $u \in V(G)$. Then*

$$\overline{IP}(G) \;\geqslant\; IP(G),$$

with equality if and only if G is $\frac{\mathfrak{n}-1}{2}$-regular (in particular, this requires $\mathfrak{n}$ odd).

Proof. Using the vertex form,

$$IP(G) = \sum_{u \in V} \sqrt{\delta_G(u)}, \qquad \overline{IP}(G) = \sum_{u \in V} \frac{\mathfrak{n}-1-\delta_G(u)}{\sqrt{\delta_G(u)}}.$$

If $\delta_G(u) \leqslant \frac{\mathfrak{n}-1}{2}$, then $\mathfrak{n}-1-\delta_G(u) \geqslant \delta_G(u)$, hence

$$\frac{\mathfrak{n}-1-\delta_G(u)}{\sqrt{\delta_G(u)}} \geqslant \sqrt{\delta_G(u)}.$$

Summing over u yields $\overline{IP}(G) \geqslant IP(G)$. Equality holds iff $\mathfrak{n}-1-\delta_G(u) = \delta_G(u)$ for all u, i.e. $\delta_G(u) = \frac{\mathfrak{n}-1}{2}$ for all u. ■

Corollary. *Let G be a simple graph on $\mathfrak{n} \geqslant 2$ vertices such that $\delta_G(u) \geqslant \left\lceil \frac{\mathfrak{n}-1}{2} \right\rceil$ for all $u \in V(G)$. Then*

$$\overline{IP}(G) \;\leqslant\; IP(G),$$

with equality if and only if G is $\frac{\mathfrak{n}-1}{2}$-regular (in particular, this requires $\mathfrak{n}$ odd). In the even-$\mathfrak{n}$ case, the inequality is strict.

Corollary. *Let $G = (V, E)$ be a simple graph on $\mathfrak{n} \geqslant 2$ vertices, with $\mathfrak{m} = |E|$ and complement $\overline{G}$ having $\overline{\mathfrak{m}}$ edges. Then*

$$\frac{2\mathfrak{m}}{\sqrt{\Delta(G)}} + \frac{2\overline{\mathfrak{m}}}{\sqrt{\Delta(G)}} \;\leqslant\; IP(G) + \overline{IP}(G) \;\leqslant\; \frac{2\mathfrak{m}}{\sqrt{\delta(G)}} + \frac{2\overline{\mathfrak{m}}}{\sqrt{\delta(G)}},$$

with equality iff G is regular.

Proof. Combine Theorem 14 with Theorem 15. ■

Theorem 18. *For any simple graph G on $\mathfrak{n} \geqslant 2$ vertices,*

$$IP(G) \cdot \overline{IP}(G) \;\leqslant\; \tfrac{1}{2}\,\mathfrak{n}^2(\mathfrak{n}-1).$$

Equality holds if and only if G is $\frac{\mathfrak{n}-1}{2}$-regular (so $\overline{G}$ is also regular) and $\mathfrak{m} = \overline{\mathfrak{m}}$, which is possible, for example, when $\mathfrak{n} \equiv 1 \pmod 4$.

ARTICLE IN PRESS

ARTICLE IN PRESS

Proof. By Theorem 13, $IP(G) \leqslant \sqrt{2\mathfrak{n}\,\mathfrak{m}}$. Also, by applying Cauchy-Schwarz to the vertex form

$$\overline{IP}(G) = \sum_{u\in V} \frac{\mathfrak{n}-1-\delta_G(u)}{\sqrt{\delta_G(u)}},$$

we have the coarse bound $\overline{IP}(G) \leqslant 2\overline{\mathfrak{m}}$, and hence (for a symmetric product estimate) we use $\overline{IP}(G) \leqslant \sqrt{2\mathfrak{n}\,\overline{\mathfrak{m}}}$ (which follows by applying Cauchy-Schwarz to the edge-form definition of $\overline{IP}(G)$). Therefore,

$$IP(G)\cdot\overline{IP}(G) \leqslant 2\mathfrak{n}\sqrt{\mathfrak{m}\,\overline{\mathfrak{m}}} \leqslant 2\mathfrak{n}\cdot\frac{\mathfrak{m}+\overline{\mathfrak{m}}}{2} = \tfrac{1}{2}\,\mathfrak{n}^2(\mathfrak{n}-1),$$

since $\mathfrak{m}+\overline{\mathfrak{m}} = \binom{\mathfrak{n}}{2}$. Equality occurs precisely when G and $\overline{G}$ are regular and $\mathfrak{m} = \overline{\mathfrak{m}}$, i.e. when G is $(\mathfrak{n}-1)/2$-regular. ■

The following results compare the Inverse Prodeg index with several well-known degree-based descriptors, including the first inverse Nirmala index, sum-connectivity, a Randić-type quantity, and the misbalance prodeg index.

Theorem 19. *Let $G = (V, E)$ be a simple graph with $E \neq \varnothing$ and $\delta(G) \geqslant 1$. Then*

$$IN_1(G) < IP(G),$$

where

$$IN_1(G) = \sum_{uv\in E(G)} \sqrt{\frac{1}{\delta_G(u)} + \frac{1}{\delta_G(v)}}.$$

Proof. Fix an edge $uv \in E(G)$ and set $x = \delta_G(u)$, $y = \delta_G(v)$, where $x, y \in \mathbb{Z}_{>0}$ (since $\delta(G) \geqslant 1$). Define the two endpoint weights

$$A(x,y) = \frac{1}{\sqrt{x}} + \frac{1}{\sqrt{y}}, \qquad B(x,y) = \sqrt{\frac{1}{x} + \frac{1}{y}}.$$

Then

$$B(x,y) = \frac{\sqrt{x+y}}{\sqrt{xy}}, \qquad A(x,y) = \frac{\sqrt{x}+\sqrt{y}}{\sqrt{xy}}.$$

Since $\sqrt{xy} > 0$, the comparison reduces to $\sqrt{x+y} < \sqrt{x} + \sqrt{y}$, which follows from

$$(\sqrt{x}+\sqrt{y})^2 = x + y + 2\sqrt{xy} > x + y = (\sqrt{x+y})^2.$$

Thus $B(x,y) < A(x,y)$ for every edge uv. Summing over $E(G)$ gives $IN_1(G) < IP(G)$. ■

Corollary. *Let $G = (V, E)$ be a simple graph with $\delta(G) \geqslant 1$. Then*

$$\chi(G) < IP(G),$$

where we use the sum-connectivity index $\chi(G) = \sum_{uv\in E(G)} \big(\delta_G(u) + \delta_G(v)\big)^{-1/2}$.

Proof. For $x, y > 0$, since $x < x + y$ and $y < x + y$, we have

$$\frac{1}{\sqrt{x}} > \frac{1}{\sqrt{x+y}}, \qquad \frac{1}{\sqrt{y}} > 0.$$

Therefore

$$\frac{1}{\sqrt{x+y}} < \frac{1}{\sqrt{x}} + \frac{1}{\sqrt{y}}.$$

Applying this to $x = \delta_G(u)$, $y = \delta_G(v)$, and summing over $uv \in E(G)$, gives $\chi(G) < IP(G)$. ■

Corollary. *Let G be an r-regular simple graph on $\mathfrak{n}$ vertices (so $\mathfrak{m} = \frac{\mathfrak{n}r}{2}$ and $\overline{\mathfrak{m}} = \frac{\mathfrak{n}(\mathfrak{n}-1-r)}{2}$). Then*

$$IP(G) = \mathfrak{n}\sqrt{r}.$$

Equivalently,

$$IP(G) = \frac{2\mathfrak{m}}{\sqrt{r}}.$$

ARTICLE IN PRESS

Moreover, for the coindex (defined using degrees in G)

$$\overline{IP}(G) = \overline{\mathfrak{m}} \cdot \frac{2}{\sqrt{r}} = \frac{2\overline{\mathfrak{m}}}{\sqrt{r}} = \frac{\mathfrak{n}(\mathfrak{n}-1-r)}{\sqrt{r}}.$$

Consequently,

$$IP(G) \leqslant \overline{IP}(G) \iff r \leqslant \frac{\mathfrak{n}-1}{2},$$

with equality iff $r = \frac{\mathfrak{n}-1}{2}$ (hence only when $\mathfrak{n}$ is odd).

Theorem 20. *Let $G = (V, E)$ be a simple graph with $\delta(G) \geqslant 1$. Then:*

$$IP(G) \geqslant 2\, R^{-\frac{1}{4}}(G), \tag{4}$$

$$IP(G) \geqslant \frac{4\mathfrak{m}^2}{MPI(G)}, \tag{5}$$

where

$$R^{-\frac{1}{4}}(G) = \sum_{uv \in E(G)} \big(\delta_G(u)\delta_G(v)\big)^{-1/4}, \qquad MPI(G) = \sum_{u \in V(G)} \delta_G(u)^{3/2}.$$

Equality in (4) *and* (5) *holds if and only if G is regular.*

Proof. Proof of (4). By AM-GM, for each $uv \in E(G)$,

$$\frac{1}{\sqrt{\delta_G(u)}} + \frac{1}{\sqrt{\delta_G(v)}} \geqslant 2\sqrt{\frac{1}{\sqrt{\delta_G(u)}} \cdot \frac{1}{\sqrt{\delta_G(v)}}} = 2\big(\delta_G(u)\delta_G(v)\big)^{-1/4}.$$

Summing over $E(G)$ yields (4). Equality holds iff $\delta_G(u) = \delta_G(v)$ for every edge uv, which forces regularity.

Proof of (5). Using $IP(G) = \sum_{u \in V} \delta_G(u)^{1/2}$ and Cauchy-Schwarz.

$$\left(\sum_{u \in V} \delta_G(u)^{1/2}\right)\left(\sum_{u \in V} \delta_G(u)^{3/2}\right) \geqslant \left(\sum_{u \in V} \delta_G(u)\right)^2 = (2\mathfrak{m})^2.$$

Rearranging gives (5). Equality holds iff $\delta_G(u)^{1/2}$ is proportional to $\delta_G(u)^{3/2}$ across V, i.e. iff $\delta_G(u)$ is constant, hence G is regular. ■

ARTICLE IN PRESS

4 Bounds for Inverse Prodeg Index of Graph Operations

4.1 Cartesian product

The Cartesian product of G_1 and G_2, denoted $G_1 \square G_2$, has vertex set $V(G_1) \times V(G_2)$. Two vertices (u_a, v_t) and (u_b, v_s) are adjacent if either

- $u_a = u_b$ and $v_t v_s \in E(G_2)$, or
- $v_t = v_s$ and $u_a u_b \in E(G_1)$.

The degree of $(u, v) \in V(G_1 \square G_2)$ is

$$\delta_{G_1 \square G_2}(u, v) = \delta_{G_1}(u) + \delta_{G_2}(v).$$

We use the well-known identity $\delta_{G_1 \square \cdots \square G_k}(v_1, \ldots, v_k) = \sum_{i=1}^{k} \delta_{G_i}(v_i)$.

Theorem 21. *Let $G_1, \ldots, G_k$ be simple graphs with $\mathfrak{n}_i = |V(G_i)|$ and $\mathfrak{m}_i = |E(G_i)|$, and set*

$$G = \square_{i=1}^{k} G_i.$$

Then

$$IP(G) = \sum_{\mathbf{v}=(v_1,\ldots,v_k) \in V(G)} \sqrt{\sum_{i=1}^{k} \delta_{G_i}(v_i)}.$$

ARTICLE IN PRESS

Moreover,

$$IP\big(\square_{i=1}^{k} G_i\big) \;\leqslant\; \Big(\prod_{i=1}^{k} \mathfrak{n}_i\Big)\sqrt{\sum_{i=1}^{k} \frac{2\mathfrak{m}_i}{\mathfrak{n}_i}}\,,$$

with equality if and only if each G_i is regular; and

$$IP\big(\square_{i=1}^{k} G_i\big) \;\geqslant\; \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Big(\prod_{j\neq i} \mathfrak{n}_j\Big)\, IP(G_i),$$

with equality if and only if either $k = 1$ or (for $k \geqslant 2$) all G_i are r-regular with the same degree r.

Proof. For $\mathbf{v} = (v_1, \ldots, v_k)$,

$$\delta_G(\mathbf{v}) = \sum_{i=1}^{k} \delta_{G_i}(v_i).$$

Upper bound. By Cauchy-Schwarz (Theorem 13),

$$IP(G) = \sum_{\mathbf{v}\in V(G)} \sqrt{\delta_G(\mathbf{v})} \leqslant \sqrt{|V(G)| \sum_{\mathbf{v}\in V(G)} \delta_G(\mathbf{v})} = \sqrt{|V(G)| \cdot 2|E(G)|}.$$

Now $|V(G)| = \prod_{i=1}^{k} \mathfrak{n}_i$. Also, the average degree of the Cartesian product equals the sum of the average degrees of the factors, hence

$$\frac{2|E(G)|}{|V(G)|} = \sum_{i=1}^{k} \frac{2\mathfrak{m}_i}{\mathfrak{n}_i} \qquad \Longrightarrow \qquad 2|E(G)| = \Big(\prod_{i=1}^{k} \mathfrak{n}_i\Big) \sum_{i=1}^{k} \frac{2\mathfrak{m}_i}{\mathfrak{n}_i}.$$

Substituting yields the desired upper bound. Equality in Cauchy-Schwarz holds iff $\delta_G(\mathbf{v})$ is constant over $V(G)$, which is equivalent to each G_i being regular.

Lower bound. For nonnegative reals $x_1, \ldots, x_k$,

$$\sqrt{\sum_{i=1}^{k} x_i} \geqslant \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \sqrt{x_i} \qquad \text{(by Cauchy-Schwarz)}.$$

Taking $x_i = \delta_{G_i}(v_i)$ and summing over all $\mathbf{v} = (v_1, \ldots, v_k)$ gives

$$IP(G) \geqslant \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \sum_{\mathbf{v}\in V(G)} \sqrt{\delta_{G_i}(v_i)} = \frac{1}{\sqrt{k}} \sum_{i=1}^{k} \Big(\prod_{j\neq i} \mathfrak{n}_j\Big)\, IP(G_i).$$

Equality in the scalar inequality holds iff $\sqrt{\delta_{G_1}(v_1)} = \cdots = \sqrt{\delta_{G_k}(v_k)}$ for every $\mathbf{v}$, which occurs precisely when each G_i is regular and all degrees are equal (or trivially when $k = 1$). ■

Corollary. *The exact values and bounds for the IP-index of some special Cartesian products are given in Table 3 below, using P_p for a path on p vertices and C_q for a cycle on q vertices.*

Graphs	IP **exact value**	IP **upper bound**	IP **lower bound**
Torus $C_{q_1} \square \cdots \square C_{q_k}$	$\Big(\prod_{i=1}^{k} q_i\Big)\sqrt{2k}$	$\Big(\prod_{i=1}^{k} q_i\Big)\sqrt{2k}$	$\Big(\prod_{i=1}^{k} q_i\Big)\sqrt{2k}$
Hamming graph $K_{\mathfrak{n}_1} \square \cdots \square K_{\mathfrak{n}_k}$	$\Big(\prod_{i=1}^{k} \mathfrak{n}_i\Big)\sqrt{\sum_{i=1}^{k}(\mathfrak{n}_i - 1)}$	$\Big(\prod_{i=1}^{k} \mathfrak{n}_i\Big)\sqrt{\sum_{i=1}^{k}(\mathfrak{n}_i - 1)}$	$\frac{1}{\sqrt{k}}\Big(\prod_{j=1}^{k} \mathfrak{n}_j\Big)\sum_{i=1}^{k}\sqrt{\mathfrak{n}_i - 1}$
k-dimensional hypercube $Q_k = \underbrace{K_2 \square \cdots \square K_2}_{k \text{ times}}$	$2^k\sqrt{k}$	$2^k\sqrt{k}$	$2^k\sqrt{k}$
$K_{\mathfrak{n}_1} \square K_{\mathfrak{n}_2}$ (rook/Hamming grid)	$\mathfrak{n}_1\mathfrak{n}_2\sqrt{(\mathfrak{n}_1 - 1) + (\mathfrak{n}_2 - 1)}$	$\mathfrak{n}_1\mathfrak{n}_2\sqrt{(\mathfrak{n}_1 - 1) + (\mathfrak{n}_2 - 1)}$	$\frac{\mathfrak{n}_1\mathfrak{n}_2}{\sqrt{2}}\big(\sqrt{\mathfrak{n}_1 - 1} + \sqrt{\mathfrak{n}_2 - 1}\big)$
Nanotube $R = P_p \square C_q$	-	$pq\sqrt{\big(\frac{2(p-1)}{p}\big) + 2}$	$\frac{1}{\sqrt{2}}\Big[\, q\, IP(P_p) \;+\; p\, IP(C_q)\,\Big]$
Nanotorus $S = C_q \square C_r$	$2qr$	$2qr$	$2qr$

Table 3. Exact values and bounds for the Inverse Prodeg index IP under Cartesian products.

ARTICLE IN PRESS

4.2 Strong product

Bounds for Inverse Prodeg Index of Strong Product. The strong product of two graphs G_1 and G_2 is denoted by $G_1 \boxtimes G_2$ and has vertex set $V(G_1) \times V(G_2)$. Two vertices (u_1, v_1) and (u_2, v_2) are adjacent in $G_1 \boxtimes G_2$ if and only if

$$\big[u_1 = u_2 \ \wedge \ v_1v_2 \in E(G_2)\big] \ \vee \ \big[v_1 = v_2 \ \wedge \ u_1u_2 \in E(G_1)\big] \ \vee \ \big[u_1u_2 \in E(G_1) \ \wedge \ v_1v_2 \in E(G_2)\big].$$

The degree of $(u, v) \in V(G_1 \boxtimes G_2)$ is

$$\delta_{G_1\boxtimes G_2}(u, v) = \delta_{G_1}(u) + \delta_{G_2}(v) + \delta_{G_1}(u)\,\delta_{G_2}(v).$$

Theorem 22. *Let G_1, G_2 be simple graphs with $\mathfrak{n}_i = |V(G_i)|$, $\mathfrak{m}_i = |E(G_i)|$, and $\delta_i = \delta(G_i)$ $(i = 1, 2)$. Then*

$$\mathfrak{n}_1\mathfrak{n}_2\sqrt{\delta_1 + \delta_2 + \delta_1\delta_2} \ \leqslant \ IP(G_1 \boxtimes G_2) \ \leqslant \ \mathfrak{n}_1\mathfrak{n}_2\sqrt{\frac{2\mathfrak{m}_1}{\mathfrak{n}_1} + \frac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \frac{4\mathfrak{m}_1\mathfrak{m}_2}{\mathfrak{n}_1\mathfrak{n}_2}}.$$

Moreover, equality holds in either (hence both) bounds if and only if G_1 and G_2 are regular.

Proof. Lower bound. For all $u \in V(G_1)$ and $v \in V(G_2)$ we have $\delta_{G_1}(u) \geqslant \delta_1$ and $\delta_{G_2}(v) \geqslant \delta_2$, hence

$$\delta_{G_1\boxtimes G_2}(u, v) = \delta_{G_1}(u) + \delta_{G_2}(v) + \delta_{G_1}(u)\delta_{G_2}(v) \geqslant \delta_1 + \delta_2 + \delta_1\delta_2.$$

Since $x \mapsto \sqrt{x}$ is increasing,

$$IP(G_1 \boxtimes G_2) = \sum_{(u,v)\in V(G_1)\times V(G_2)} \sqrt{\delta_{G_1\boxtimes G_2}(u, v)} \geqslant \mathfrak{n}_1\mathfrak{n}_2\sqrt{\delta_1 + \delta_2 + \delta_1\delta_2}.$$

Equality forces $\delta_{G_1}(u)$ and $\delta_{G_2}(v)$ to be constant, hence G_1 and G_2 are regular.

Upper bound. Write

$$IP(G_1 \boxtimes G_2) = \sum_{u\in V(G_1)} \sum_{v\in V(G_2)} \sqrt{\delta_{G_1}(u) + \delta_{G_2}(v) + \delta_{G_1}(u)\delta_{G_2}(v)}.$$

Fix $u \in V(G_1)$ and define

$$x_v = \delta_{G_1}(u) + \delta_{G_2}(v) + \delta_{G_1}(u)\delta_{G_2}(v) \qquad (v \in V(G_2)).$$

Since $f(x) = \sqrt{x}$ is concave on $(0, \infty)$, Jensen's inequality yields

$$\frac{1}{\mathfrak{n}_2} \sum_{v\in V(G_2)} \sqrt{x_v} \leqslant \sqrt{\frac{1}{\mathfrak{n}_2} \sum_{v\in V(G_2)} x_v}.$$

Compute

$$\sum_{v\in V(G_2)} x_v = \mathfrak{n}_2\,\delta_{G_1}(u) + \sum_v \delta_{G_2}(v) + \delta_{G_1}(u) \sum_v \delta_{G_2}(v) = \mathfrak{n}_2\,\delta_{G_1}(u) + 2\mathfrak{m}_2 + 2\mathfrak{m}_2\,\delta_{G_1}(u),$$

so

$$\sum_{v\in V(G_2)} \sqrt{x_v} \leqslant \mathfrak{n}_2\sqrt{\delta_{G_1}(u) + \frac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \delta_{G_1}(u)\frac{2\mathfrak{m}_2}{\mathfrak{n}_2}}.$$

Now sum over $u \in V(G_1)$:

$$IP(G_1 \boxtimes G_2) \leqslant \mathfrak{n}_2 \sum_{u\in V(G_1)} \sqrt{\delta_{G_1}(u) + \frac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \delta_{G_1}(u)\frac{2\mathfrak{m}_2}{\mathfrak{n}_2}}.$$

Apply Jensen again over u to the explicitly defined concave function $f(t) = \sqrt{t}$, $t > 0$:

$$\frac{1}{\mathfrak{n}_1} \sum_{u\in V(G_1)} \sqrt{y_u} \leqslant \sqrt{\frac{1}{\mathfrak{n}_1} \sum_{u\in V(G_1)} y_u},$$

where

$$y_u = \delta_{G_1}(u) + \frac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \delta_{G_1}(u)\frac{2\mathfrak{m}_2}{\mathfrak{n}_2}.$$

ARTICLE IN PRESS

Since $\sum_u \delta_{G_1}(u) = 2\mathfrak{m}_1$, we obtain

$$\frac{1}{\mathfrak{n}_1} \sum_{u \in V(G_1)} y_u = \frac{2\mathfrak{m}_1}{\mathfrak{n}_1} + \frac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \frac{4\mathfrak{m}_1\mathfrak{m}_2}{\mathfrak{n}_1\mathfrak{n}_2}.$$

Therefore,

$$IP(G_1 \boxtimes G_2) \leqslant \mathfrak{n}_1\mathfrak{n}_2 \sqrt{\frac{2\mathfrak{m}_1}{\mathfrak{n}_1} + \frac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \frac{4\mathfrak{m}_1\mathfrak{m}_2}{\mathfrak{n}_1\mathfrak{n}_2}}.$$

Equality in both Jensen steps forces x_v to be constant in v for each fixed u and y_u to be constant in u, which is equivalent to $\delta_{G_2}(v)$ being constant on $V(G_2)$ and $\delta_{G_1}(u)$ being constant on $V(G_1)$; hence G_1 and G_2 are regular. ■

4.3 Disjunction

Bounds for Inverse Prodeg Index of Disjunction. The disjunction $G_1 \vee G_2$ has vertex set $V(G_1) \times V(G_2)$ with $(u_1, v_1) \sim (u_2, v_2)$ if and only if $u_1u_2 \in E(G_1)$ or $v_1v_2 \in E(G_2)$. Its degree is

$$\delta_{G_1 \vee G_2}(u_1, v_1) = \mathfrak{n}_2\, \delta_{G_1}(u_1) + \mathfrak{n}_1\, \delta_{G_2}(v_1) - \delta_{G_1}(u_1)\, \delta_{G_2}(v_1).$$

Theorem 23. *Let G_1, G_2 have $\mathfrak{n}_i = |V(G_i)|$, $\mathfrak{m}_i = |E(G_i)|$, minimum and maximum degrees δ_i, Δ_i. Define*

$$\delta_{\min} = \min_{(u,v) \in V(G_1) \times V(G_2)} \delta_{G_1 \vee G_2}(u, v), \qquad \Delta_{\max} = \max_{(u,v) \in V(G_1) \times V(G_2)} \delta_{G_1 \vee G_2}(u, v).$$

Then

$$\text{(Lower)} \qquad \mathfrak{n}_1\mathfrak{n}_2 \sqrt{\delta_{\min}} \;\leqslant\; IP(G_1 \vee G_2),$$

$$\text{(Upper 1)} \qquad IP(G_1 \vee G_2) \;\leqslant\; \mathfrak{n}_1\mathfrak{n}_2 \sqrt{\Delta_{\max}},$$

$$\text{(Upper 2)} \qquad IP(G_1 \vee G_2) \;\leqslant\; \mathfrak{n}_2^{3/2}\, IP(G_1) \;+\; \mathfrak{n}_1^{3/2}\, IP(G_2).$$

Equality in (Lower) *and* (Upper 1) *holds iff G_1 and G_2 are regular (equivalently, $G_1 \vee G_2$ is regular). Equality in* (Upper 2) *holds iff one factor is edgeless (i.e. $\mathfrak{m}_1 = 0$ or $\mathfrak{m}_2 = 0$).*

Proof. Lower/Upper 1. From $\delta_{\min} \leqslant \delta_{G_1 \vee G_2}(u, v) \leqslant \Delta_{\max}$ and monotonicity of $\sqrt{\cdot}$,

$$\mathfrak{n}_1\mathfrak{n}_2 \sqrt{\delta_{\min}} \leqslant \sum_{(u,v) \in V(G_1) \times V(G_2)} \sqrt{\delta_{G_1 \vee G_2}(u, v)} \leqslant \mathfrak{n}_1\mathfrak{n}_2 \sqrt{\Delta_{\max}}.$$

Both inequalities are equalities iff $\delta_{G_1 \vee G_2}$ is constant on $V(G_1) \times V(G_2)$, which occurs exactly when G_1 and G_2 are regular.

Upper 2. Since $-\delta_{G_1}(u)\delta_{G_2}(v) \leqslant 0$, we have

$$\delta_{G_1 \vee G_2}(u, v) \leqslant \mathfrak{n}_2 \delta_{G_1}(u) + \mathfrak{n}_1 \delta_{G_2}(v).$$

Using $\sqrt{x + y} \leqslant \sqrt{x} + \sqrt{y}$ for $x, y \geqslant 0$,

$$\sqrt{\delta_{G_1 \vee G_2}(u, v)} \leqslant \sqrt{\mathfrak{n}_2 \delta_{G_1}(u)} + \sqrt{\mathfrak{n}_1 \delta_{G_2}(v)}.$$

Summing over $V(G_1) \times V(G_2)$ gives

$$IP(G_1 \vee G_2) \leqslant \mathfrak{n}_2^{3/2} \sum_{u \in V(G_1)} \sqrt{\delta_{G_1}(u)} + \mathfrak{n}_1^{3/2} \sum_{v \in V(G_2)} \sqrt{\delta_{G_2}(v)} = \mathfrak{n}_2^{3/2} IP(G_1) + \mathfrak{n}_1^{3/2} IP(G_2).$$

If equality holds in *(Upper 2)*, then for every (u, v) we must have $\delta_{G_1}(u)\delta_{G_2}(v) = 0$, forcing $\delta_{G_1} \equiv 0$ or $\delta_{G_2} \equiv 0$, i.e. $\mathfrak{m}_1 = 0$ or $\mathfrak{m}_2 = 0$. ■

4.4 Symmetric difference

Bounds for Inverse Prodeg Index of Symmetric Difference. The symmetric difference $G_1 \oplus G_2$ has vertex set $V(G_1) \times V(G_2)$, where $(u_1, v_1) \sim (u_2, v_2)$ if exactly one of $u_1u_2 \in E(G_1)$ or $v_1v_2 \in E(G_2)$ holds. Equivalently, the

ARTICLE IN PRESS

adjacency condition is an exclusive-or condition: XOR means that exactly one, but not both, of the two stated edge relations is true. Its degree satisfies

$$\delta_{G_1\oplus G_2}(u,v) = \mathfrak{n}_2\delta_{G_1}(u) + \mathfrak{n}_1\delta_{G_2}(v) - 2\delta_{G_1}(u)\,\delta_{G_2}(v).$$

Theorem 24. *Let G_1, G_2 be simple graphs with $\mathfrak{n}_i = |V(G_i)|$, and minimum and maximum degrees $\delta_i = \delta(G_i)$ and $\Delta_i = \Delta(G_i)$ $(i = 1, 2)$. Define*

$$\delta_{\min} = \min_{(u,v)\in V(G_1)\times V(G_2)} \delta_{G_1\oplus G_2}(u,v), \qquad \Delta_{\max} = \max_{(u,v)\in V(G_1)\times V(G_2)} \delta_{G_1\oplus G_2}(u,v).$$

Then

$$\mathfrak{n}_1\mathfrak{n}_2\sqrt{\delta_{\min}} \;\leqslant\; IP(G_1\oplus G_2) \;\leqslant\; \mathfrak{n}_1\mathfrak{n}_2\sqrt{\Delta_{\max}}.$$

Moreover, equality holds in either inequality if and only if $G_1 \oplus G_2$ is regular.

Proof. By definition,

$$IP(G_1\oplus G_2) = \sum_{(u,v)\in V(G_1)\times V(G_2)} \sqrt{\delta_{G_1\oplus G_2}(u,v)}.$$

Since $\delta_{\min} \leqslant \delta_{G_1\oplus G_2}(u,v) \leqslant \Delta_{\max}$ for all (u,v) and $\sqrt{\cdot}$ is increasing, we obtain

$$\mathfrak{n}_1\mathfrak{n}_2\sqrt{\delta_{\min}} \leqslant \sum_{(u,v)} \sqrt{\delta_{G_1\oplus G_2}(u,v)} \leqslant \mathfrak{n}_1\mathfrak{n}_2\sqrt{\Delta_{\max}}.$$

Equality in either inequality holds if and only if $\delta_{G_1\oplus G_2}(u,v)$ is constant over $V(G_1)\times V(G_2)$, i.e. iff $G_1\oplus G_2$ is regular. ■

Additional conditional upper bounds.

Proposition 25. *Let G_1, G_2 be simple graphs on $\mathfrak{n}_1, \mathfrak{n}_2$ vertices.*

(A) Edgeless-factor reduction. *If $\mathfrak{m}_1 = 0$ (i.e. G_1 is edgeless), then*

$$G_1 \oplus G_2 \cong \overline{G_1 \vee G_2},$$

and if $\mathfrak{m}_2 = 0$ then

$$G_1 \oplus G_2 \cong \overline{G_1 \vee G_2}.$$

In particular, the bounds from Theorem 23 can be transferred to $G_1 \oplus G_2$ by applying them to $G_1 \vee G_2$ and then taking complements if needed.

(B) A uniform degree-restriction bound. *Assume that for all $(u,v) \in V(G_1)\times V(G_2)$,*

$$\mathfrak{n}_2\,\delta_{G_1}(u) + \mathfrak{n}_1\,\delta_{G_2}(v) \;\geqslant\; 2\,\delta_{G_1}(u)\,\delta_{G_2}(v). \tag{6}$$

Then

$$IP(G_1\oplus G_2) \;\leqslant\; \mathfrak{n}_2^{3/2}\,IP(G_1) \;+\; \mathfrak{n}_1^{3/2}\,IP(G_2).$$

Proof. (A) Suppose $\mathfrak{m}_1 = 0$. Then $E(G_1) = \varnothing$, so for any two distinct vertices $(u_1,v_1),(u_2,v_2)$ we have

$$(u_1,v_1) \sim_{G_1\oplus G_2} (u_2,v_2) \iff \big[u_1u_2 \in E(G_1)\big] \text{ XOR } \big[v_1v_2 \in E(G_2)\big] \iff v_1v_2 \in E(G_2).$$

Hence the adjacency in $G_1 \oplus G_2$ occurs exactly when $v_1v_2 \in E(G_2)$ and is independent of u_1, u_2, while in the disjunction $G_1 \vee G_2$ (with $E(G_1) = \varnothing$) adjacency occurs when $v_1v_2 \in E(G_2)$ *or* $u_1 \neq u_2$; thus $G_1 \vee G_2$ is obtained from $G_1 \oplus G_2$ by adding all edges between different u-layers, i.e. $G_1 \oplus G_2$ is the complement of $G_1 \vee G_2$ in $V(G_1)\times V(G_2)$. This gives $G_1 \oplus G_2 \cong \overline{G_1 \vee G_2}$, and the case $\mathfrak{m}_2 = 0$ is analogous.

(B) By the degree formula,

$$\delta_{G_1\oplus G_2}(u,v) = \mathfrak{n}_2\,\delta_{G_1}(u) + \mathfrak{n}_1\,\delta_{G_2}(v) - 2\,\delta_{G_1}(u)\delta_{G_2}(v).$$

ARTICLE IN PRESS

Under (6), the right-hand side is nonnegative and satisfies

$$\delta_{G_1\oplus G_2}(u,v) \leqslant \mathfrak{n}_2\,\delta_{G_1}(u) + \mathfrak{n}_1\,\delta_{G_2}(v).$$

Therefore, using $\sqrt{x+y} \leqslant \sqrt{x} + \sqrt{y}$ for $x, y \geqslant 0$,

$$\sqrt{\delta_{G_1\oplus G_2}(u,v)} \leqslant \sqrt{\mathfrak{n}_2\,\delta_{G_1}(u) + \mathfrak{n}_1\,\delta_{G_2}(v)} \leqslant \sqrt{\mathfrak{n}_2\,\delta_{G_1}(u)} + \sqrt{\mathfrak{n}_1\,\delta_{G_2}(v)}.$$

Summing over $V(G_1) \times V(G_2)$ yields

$$IP(G_1 \oplus G_2) \leqslant \mathfrak{n}_2^{3/2} \sum_{u\in V(G_1)} \sqrt{\delta_{G_1}(u)} + \mathfrak{n}_1^{3/2} \sum_{v\in V(G_2)} \sqrt{\delta_{G_2}(v)} = \mathfrak{n}_2^{3/2} IP(G_1) + \mathfrak{n}_1^{3/2} IP(G_2).$$

■

Remark. A simple sufficient condition ensuring (6) is

$$\Delta(G_1) \leqslant \frac{\mathfrak{n}_1}{2} \quad \text{and} \quad \Delta(G_2) \leqslant \frac{\mathfrak{n}_2}{2}. \tag{7}$$

Indeed, for any $u \in V(G_1)$ and $v \in V(G_2)$ we have

$$\mathfrak{n}_2\,\delta_{G_1}(u) + \mathfrak{n}_1\,\delta_{G_2}(v) - 2\,\delta_{G_1}(u)\delta_{G_2}(v) \;\geqslant\; \mathfrak{n}_2\,\delta_{G_1}(u) - 2\,\delta_{G_1}(u)\Delta(G_2) = \delta_{G_1}(u)\big(\mathfrak{n}_2 - 2\Delta(G_2)\big) \geqslant 0,$$

and similarly

$$\mathfrak{n}_2\,\delta_{G_1}(u) + \mathfrak{n}_1\,\delta_{G_2}(v) - 2\,\delta_{G_1}(u)\delta_{G_2}(v) \;\geqslant\; \delta_{G_2}(v)\big(\mathfrak{n}_1 - 2\Delta(G_1)\big) \geqslant 0.$$

Hence (6) holds for all (u, v) whenever (7) is satisfied.

4.5 Tensor product

***IP*-index under Tensor Product.** The 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 $v_1v_2 \in E(G_2)$. For $(u, v) \in V(G_1) \times V(G_2)$,

$$\delta_{G_1\otimes G_2}(u,v) = \delta_{G_1}(u)\,\delta_{G_2}(v).$$

Theorem 26. *For any finite simple graphs G_1 and G_2,*

$$IP(G_1 \otimes G_2) = IP(G_1) \cdot IP(G_2).$$

Proof. Using $\delta_{G_1\otimes G_2}(u,v) = \delta_{G_1}(u)\delta_{G_2}(v)$, we have

$$\begin{aligned}
IP(G_1 \otimes G_2) &= \sum_{(u,v)\in V(G_1)\times V(G_2)} \sqrt{\delta_{G_1\otimes G_2}(u,v)} \\
&= \sum_{u\in V(G_1)} \sum_{v\in V(G_2)} \sqrt{\delta_{G_1}(u)\,\delta_{G_2}(v)} \\
&= \sum_{u\in V(G_1)} \sum_{v\in V(G_2)} \sqrt{\delta_{G_1}(u)}\,\sqrt{\delta_{G_2}(v)} \\
&= \left(\sum_{u\in V(G_1)} \sqrt{\delta_{G_1}(u)}\right)\left(\sum_{v\in V(G_2)} \sqrt{\delta_{G_2}(v)}\right) \\
&= IP(G_1) \cdot IP(G_2).
\end{aligned}$$

■

Table 4 lists IP for tensor products of several standard graphs, obtained directly from Theorem 26 and the closed forms in Section 2.

ARTICLE IN PRESS

Table 4. *IP*-index for tensor products of common graphs

Graph G_1	**Graph** G_2	$\boldsymbol{IP(G_1 \otimes G_2)}$
$K_{\mathfrak{m}}$	$K_{\mathfrak{n}}$	$\mathfrak{m}\sqrt{\mathfrak{m}-1}\cdot\mathfrak{n}\sqrt{\mathfrak{n}-1} = \mathfrak{m}\mathfrak{n}\sqrt{(\mathfrak{m}-1)(\mathfrak{n}-1)}$
$P_{\mathfrak{m}}$	$P_{\mathfrak{n}}$	$\Big(2+(\mathfrak{m}-2)\sqrt{2}\Big)\Big(2+(\mathfrak{n}-2)\sqrt{2}\Big)$
$C_{\mathfrak{m}}$	$C_{\mathfrak{n}}$	$\mathfrak{m}\sqrt{2}\cdot\mathfrak{n}\sqrt{2} = 2\mathfrak{m}\mathfrak{n}$
$S_{\mathfrak{m}}$	$S_{\mathfrak{n}}$	$\Big(\sqrt{\mathfrak{m}-1}+(\mathfrak{m}-1)\Big)\Big(\sqrt{\mathfrak{n}-1}+(\mathfrak{n}-1)\Big)$
$K_{\mathfrak{m}_1,\mathfrak{m}_2}$	$K_{\mathfrak{n}_1,\mathfrak{n}_2}$	$\Big(\mathfrak{m}_1\sqrt{\mathfrak{m}_2}+\mathfrak{m}_2\sqrt{\mathfrak{m}_1}\Big)\Big(\mathfrak{n}_1\sqrt{\mathfrak{n}_2}+\mathfrak{n}_2\sqrt{\mathfrak{n}_1}\Big)$
$W_{\mathfrak{m}}$	$W_{\mathfrak{n}}$	$\Big(\sqrt{\mathfrak{m}-1}+(\mathfrak{m}-1)\sqrt{3}\Big)\Big(\sqrt{\mathfrak{n}-1}+(\mathfrak{n}-1)\sqrt{3}\Big)$
Q_d	Q_k	$2^d\sqrt{d}\cdot 2^k\sqrt{k} = 2^{d+k}\sqrt{dk}$
$F_{\mathfrak{m}}$	$F_{\mathfrak{n}}$	$\Big(\sqrt{2\mathfrak{m}}+2\sqrt{2}\,\mathfrak{m}\Big)\Big(\sqrt{2\mathfrak{n}}+2\sqrt{2}\,\mathfrak{n}\Big)$
$L_{\mathfrak{m}}$	$L_{\mathfrak{n}}$	$\Big(4\sqrt{2}+(2\mathfrak{m}-4)\sqrt{3}\Big)\Big(4\sqrt{2}+(2\mathfrak{n}-4)\sqrt{3}\Big)$
$P_{\mathfrak{m}}$	$C_{\mathfrak{n}}$	$\Big(2+(\mathfrak{m}-2)\sqrt{2}\Big)\Big(\mathfrak{n}\sqrt{2}\Big)$
$P_{\mathfrak{m}}$	$K_{\mathfrak{n}}$	$\Big(2+(\mathfrak{m}-2)\sqrt{2}\Big)\Big(\mathfrak{n}\sqrt{\mathfrak{n}-1}\Big)$
$C_{\mathfrak{m}}$	$K_{\mathfrak{n}}$	$\Big(\mathfrak{m}\sqrt{2}\Big)\Big(\mathfrak{n}\sqrt{\mathfrak{n}-1}\Big)$

4.6 Join

Bounds for *IP*-index of Join. The join $H = G_1 + G_2 + \cdots + G_k$ has vertex set $V(H) = \bigcup_{i=1}^{k} V(G_i)$ and includes all edges of each G_i together with all edges between distinct parts $V(G_i)$ and $V(G_j)$ $(i \neq j)$. Let $\mathfrak{n}_i = |V(G_i)|$ and $\mathfrak{n} = \sum_{i=1}^{k} \mathfrak{n}_i$. For $v \in V(G_i)$,

$$\delta_H(v) = \delta_{G_i}(v) + (\mathfrak{n} - \mathfrak{n}_i).$$

Theorem 27. *Let $G_1, \ldots, G_k$ be simple graphs with $\mathfrak{n}_i$ vertices, $\mathfrak{m}_i$ edges, and minimum degree $\delta(G_i)$. Then*

$$\text{(Upper)} \qquad IP(H) \leqslant \sum_{i=1}^{k} \mathfrak{n}_i \left(\frac{2\mathfrak{m}_i}{\mathfrak{n}_i} + \mathfrak{n} - \mathfrak{n}_i\right)^{1/2},$$

$$\text{(Lower)} \qquad IP(H) \geqslant \sum_{i=1}^{k} \mathfrak{n}_i \left(\delta(G_i) + \mathfrak{n} - \mathfrak{n}_i\right)^{1/2}.$$

Equality holds in the upper (resp. lower) bound if and only if each G_i is regular.

Proof. Upper bound. Fix i and apply Jensen's inequality to the concave function $f(x) = \sqrt{x}$ on the multiset $\{\, \delta_{G_i}(v) + (\mathfrak{n} - \mathfrak{n}_i) : v \in V(G_i) \,\}$:

$$\sum_{v \in V(G_i)} \sqrt{\delta_{G_i}(v) + (\mathfrak{n} - \mathfrak{n}_i)} \leqslant \mathfrak{n}_i \sqrt{\frac{1}{\mathfrak{n}_i} \sum_{v \in V(G_i)} \big(\delta_{G_i}(v) + (\mathfrak{n} - \mathfrak{n}_i)\big)}.$$

Since $\sum_{v \in V(G_i)} \delta_{G_i}(v) = 2\mathfrak{m}_i$, the right-hand side equals

$$\mathfrak{n}_i \sqrt{\frac{2\mathfrak{m}_i}{\mathfrak{n}_i} + \mathfrak{n} - \mathfrak{n}_i}.$$

Summing over i gives the stated upper bound. Equality in Jensen holds iff $\delta_{G_i}(v)$ is constant on $V(G_i)$, i.e. iff G_i is regular.

Lower bound. For $v \in V(G_i)$ we have $\delta_{G_i}(v) \geqslant \delta(G_i)$, hence

$$\sqrt{\delta_{G_i}(v) + (\mathfrak{n} - \mathfrak{n}_i)} \geqslant \sqrt{\delta(G_i) + (\mathfrak{n} - \mathfrak{n}_i)}.$$

Summing over $v \in V(G_i)$ and then over i yields the lower bound. Equality holds iff $\delta_{G_i}(v) = \delta(G_i)$ for all v, i.e. iff G_i is regular. ∎

Corollary. *Let $G_1, \ldots, G_k$ be finite simple graphs with $\mathfrak{n}_i = |V(G_i)|$ and $\mathfrak{m}_i = |E(G_i)|$, and let $\mathfrak{n} = \sum_{i=1}^{k} \mathfrak{n}_i$. For the join $H = G_1 + \cdots + G_k$,*

$$IP(H) \leqslant \sqrt{2\,\mathfrak{n}\,|E(H)|} = \sqrt{\mathfrak{n}\Big(2\sum_{i=1}^{k} \mathfrak{m}_i + \mathfrak{n}^2 - \sum_{i=1}^{k} \mathfrak{n}_i^2\Big)}.$$

ARTICLE IN PRESS

Equivalently,

$$IP(G_1 + \cdots + G_k) \;\leqslant\; \mathfrak{n}\sqrt{\frac{2}{\mathfrak{n}}\sum_{i=1}^{k}\mathfrak{m}_i \;+\; \sum_{1\leqslant i<j\leqslant k}\frac{2\mathfrak{n}_i\mathfrak{n}_j}{\mathfrak{n}}}\,.$$

Equality holds if and only if the join H is regular, i.e. each G_i is regular and the quantities $r_i + (\mathfrak{n} - \mathfrak{n}_i)$ are equal for all i (where r_i is the regular degree of G_i).

Proof. By Cauchy-Schwarz on H,

$$IP(H) = \sum_{x\in V(H)}\sqrt{\delta_H(x)} \leqslant \sqrt{|V(H)|\cdot\sum_{x\in V(H)}\delta_H(x)} = \sqrt{2\,|V(H)|\,|E(H)|}.$$

Now $|V(H)| = \mathfrak{n}$ and

$$|E(H)| = \sum_{i=1}^{k}\mathfrak{m}_i \;+\; \sum_{1\leqslant i<j\leqslant k}\mathfrak{n}_i\mathfrak{n}_j = \sum_{i=1}^{k}\mathfrak{m}_i \;+\; \frac{\mathfrak{n}^2 - \sum_i \mathfrak{n}_i^2}{2},$$

giving the displayed bound. Equality in Cauchy-Schwarz holds iff $\delta_H(\cdot)$ is constant, i.e. iff H is regular. ∎

Corollary. *Let G_1, G_2 be simple graphs with $\mathfrak{n}_i = |V(G_i)|$, $\mathfrak{m}_i = |E(G_i)|$, and minimum degrees $\delta(G_i)$. Then*

$$IP(G_1 + G_2) \;\leqslant\; \mathfrak{n}_1\Big(\tfrac{2\mathfrak{m}_1}{\mathfrak{n}_1} + \mathfrak{n}_2\Big)^{1/2} \;+\; \mathfrak{n}_2\Big(\tfrac{2\mathfrak{m}_2}{\mathfrak{n}_2} + \mathfrak{n}_1\Big)^{1/2},$$

$$IP(G_1 + G_2) \;\geqslant\; \mathfrak{n}_1\big(\delta(G_1) + \mathfrak{n}_2\big)^{1/2} \;+\; \mathfrak{n}_2\big(\delta(G_2) + \mathfrak{n}_1\big)^{1/2}.$$

Equality in the upper (resp. lower) bound holds iff G_1 and G_2 are regular.

Example 1. (i) **Complete multipartite.** For $\mathfrak{m}_1, \ldots, \mathfrak{m}_k \geqslant 1$ let

$$K_{\mathfrak{m}_1,\ldots,\mathfrak{m}_k} = \overline{K}_{\mathfrak{m}_1} + \cdots + \overline{K}_{\mathfrak{m}_k}, \qquad M = \sum_{j=1}^{k}\mathfrak{m}_j.$$

Each vertex in the i-th part has degree $M - \mathfrak{m}_i$, hence

$$IP(K_{\mathfrak{m}_1,\ldots,\mathfrak{m}_k}) = \sum_{i=1}^{k}\mathfrak{m}_i\sqrt{M - \mathfrak{m}_i}\,.$$

(ii) **Wheel.** For $\mathfrak{n} \geqslant 4$, the wheel graph satisfies $W_{\mathfrak{n}} = K_1 + C_{\mathfrak{n}-1}$. The center has degree $\mathfrak{n} - 1$ and each rim vertex has degree 3, therefore

$$IP(W_{\mathfrak{n}}) = \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{3}.$$

(iii) **Fan** $F_{\mathfrak{n}} = K_1 + P_{\mathfrak{n}-1}$ (total order $\mathfrak{n} \geqslant 3$). Let c be the vertex of K_1 and let $P_{\mathfrak{n}-1}$ have endpoints $p_1, p_{\mathfrak{n}-1}$. In the join, c is adjacent to all $\mathfrak{n} - 1$ path vertices, so

$$\delta(c) = \mathfrak{n} - 1.$$

Each endpoint of the path has degree 1 in $P_{\mathfrak{n}-1}$, hence degree $1 + 1 = 2$ in $F_{\mathfrak{n}}$ after adding the edge to c. Each internal path vertex has degree 2 in $P_{\mathfrak{n}-1}$, hence degree $2 + 1 = 3$ in $F_{\mathfrak{n}}$. Therefore,

$$IP(F_{\mathfrak{n}}) = \sqrt{\mathfrak{n} - 1} + 2\sqrt{2} + (\mathfrak{n} - 3)\sqrt{3}.$$

Moreover, using Theorem 27 with $G_1 = K_1$ and $G_2 = P_{\mathfrak{n}-1}$,

$$\textit{Upper bound:}\quad IP(F_{\mathfrak{n}}) \leqslant \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{\frac{2(\mathfrak{n} - 2)}{\mathfrak{n} - 1} + 1} = \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{3 - \frac{2}{\mathfrak{n} - 1}},$$

$$\textit{Lower bound:}\quad IP(F_{\mathfrak{n}}) \geqslant \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{1 + 1} = \sqrt{\mathfrak{n} - 1} + (\mathfrak{n} - 1)\sqrt{2}.$$

(iv) **Cone.** Let $C_{\mathfrak{m},\mathfrak{n}} = C_{\mathfrak{m}} + \overline{K}_{\mathfrak{n}}$ with $\mathfrak{m} \geqslant 3$ and $\mathfrak{n} \geqslant 1$. Vertices of $C_{\mathfrak{m}}$ have degree $\mathfrak{n} + 2$ in the join, while vertices of $\overline{K}_{\mathfrak{n}}$ have degree $\mathfrak{m}$. Thus

$$IP(C_{\mathfrak{m},\mathfrak{n}}) = \mathfrak{m}\sqrt{\mathfrak{n} + 2} + \mathfrak{n}\sqrt{\mathfrak{m}}.$$

ARTICLE IN PRESS

ARTICLE IN PRESS

5 QSPR Benchmarking and External Validation via Degree-Based Prodeg Descriptors

5.1 Dataset and physicochemical endpoints

Compound set. The dataset comprises $n = 90$ unique organic compounds represented by molecular name, InChI, InChIKey, molecular formula, and canonical SMILES. The series is dominated by substituted *benzoic-acid* and *salicylic-acid* (*o*-hydroxybenzoic acid) scaffolds, including two reference drug molecules (aspirin and salicylic acid) together with a broader panel of systematically modified analogs. Structural modifications include (i) *O*-acylation and esterification (e.g., acetoxy-, propionyloxy-, and related carbonate/ester motifs), (ii) alkoxy substitution (e.g., methoxy/ethoxy/propoxy), and (iii) extensive halogenation (chloro/bromo/fluoro and fluoroacyl/perfluoroalkyl motifs), as well as a small subset of nitrogen-containing derivatives (e.g., amide/nitro/cyano substituted aromatics). Across the dataset, the heavy-atom count ranges from 10 to 31 (median 17), with 64 distinct molecular formulas, providing moderate but meaningful diversity within a coherent aromatic-carboxylate family.

This compound set was selected because aromatic carboxylates provide a chemically interpretable and structurally coherent series with systematic substitution patterns, publicly traceable Cheméo records, and sufficient variation in size, branching, halogenation, and functional-group substitution to test whether degree-derived descriptors track physicochemical trends. The restricted scope also reduces scaffold heterogeneity, making it easier to interpret the graph-theoretic signal.

Physicochemical properties and data source. Physicochemical property values were compiled from Cheméo (`https://www.chemeo.com/`) (units as reported). A total of 12 endpoints were collected: $McVol$ (McGowan's characteristic volume, mL/mol), V_c (critical volume, m^3/kmol), $\Delta_f G^\circ$ (standard Gibbs free energy of formation, kJ/mol), $\Delta_f H^\circ_{\text{gas}}$ (standard gas-phase enthalpy of formation, kJ/mol), $\Delta_{fus} H^\circ$ (standard enthalpy of fusion, kJ/mol), $\Delta_{vap} H^\circ$ (standard enthalpy of vaporization, kJ/mol), $\log_{10} WS$ (Log_{10} water solubility, mol/L), $\log P_{\text{oct/wat}}$ (octanol/water partition coefficient), P_c (critical pressure, kPa), T_{boil} (normal boiling point, K), T_c (critical temperature, K), and T_{fus} (normal melting point, K). For transparency and reproducibility, the curated dataset used in modeling (including identifiers) is provided in the Supplementary Information (retrieved on 07 Jan 2026; record URLs/IDs and any selection rules are given in SI).

Predictors (Prodeg indices). The predictor space comprised four degree-based Prodeg indices, $\{Z_{-1/2}, IP, \overline{IP}, MPI\}$, computed from the hydrogen-suppressed molecular graph of each compound (definitions in Section 1). These indices were used as the sole descriptors to assess whether degree-derived topological information can explain and predict the selected physicochemical endpoints.

Chemically, the four Prodeg descriptors encode complementary aspects of the hydrogen-suppressed molecular graph. The descriptor IP gives comparatively larger relative contribution to terminal and low-degree atoms because the increment from degree one to degree two is larger under the concave square-root scale than under convex degree powers. The coindex $\overline{IP}$ measures how the same degree information is distributed across nonadjacent atom pairs, and therefore reflects global separation of low-degree environments within a molecule. In contrast, MPI is convex in the degree and emphasizes more substituted or branched environments. This complementarity explains why the descriptor set is useful for size-related and thermodynamic endpoints, while properties strongly influenced by three-dimensional packing, hydrogen bonding, or specific electronic effects require chemically richer descriptors.

Scope of the QSPR benchmark. The empirical study is intentionally restricted to a small, chemically coherent aromatic-carboxylate series and to four degree-based Prodeg descriptors. It is therefore a controlled test of whether the proposed descriptors contain useful signal, not a claim that they outperform all established descriptor families. Comparisons with broad cheminformatics descriptor sets (e.g., constitutional, electronic, 3D, and fingerprint descriptors) and with standard topological indices such as Zagreb and Randić descriptors are identified as necessary follow-up work.

Endpoints used for the core benchmark and validation split. Although 12 properties were compiled, three endpoints ($\log P_{\text{oct/wat}}$, $\Delta_{vap} H^\circ$, and $\Delta_{fus} H^\circ$) were excluded from the *core* benchmark because degree-descriptor-only models produced weak and/or unstable generalization. During revision, the tabulated values were rechecked against the curated Cheméo worksheet. The value 33.85 for aspirin belongs to the fusion enthalpy field $\Delta_{fus} H^\circ$, whereas the

ARTICLE IN PRESS

gas-phase formation enthalpy used in the core benchmark is $\Delta_f H^\circ_{\text{gas}} = -513.64$ kJ/mol for aspirin. Accordingly, the core benchmark focused on nine endpoints: $McVol$, V_c, $\Delta_f G^\circ$, $\Delta_f H^\circ_{\text{gas}}$, $\log_{10} WS$, P_c, T_{boil}, T_c, and T_{fus}.

To ensure rigorous external validation, the full dataset ($n = 90$) was partitioned into a training set ($n = 72$) and an independent test set ($n = 18$) using the Kennard-Stone algorithm applied in standardized four-dimensional Prodeg space (80:20 split). The split was constructed using predictors only (no endpoint information); standardization here is used solely for distance computation in descriptor space. All subsequent preprocessing (including scaling), cross-validation, and hyperparameter tuning were performed *exclusively* on the training set. Models were then refit on the full training set and evaluated once on the fixed external test set. To enable consistent comparisons across endpoints and accommodate intrinsically single-output learners, we fit one independent model per endpoint (nine separate fits) under the same split.

Table 5 lists the four Prodeg descriptors and the nine core benchmark endpoints for all 90 compounds (values rounded to two decimals).

ARTICLE IN PRESS

ARTICLE IN PRESS

Abbreviations for Table 5: $McVol$ = McGowan's characteristic volume (mL/mol), V_c = Critical volume (m^3/kmol), $\Delta_f G^\circ$ = Standard Gibbs free energy of formation (kJ/mol), $\Delta_f H^\circ_{\text{gas}}$ = Standard gas-phase enthalpy of formation (kJ/mol), $\log_{10} WS$ = Log_{10} of water solubility (mol/L), P_c = Critical pressure (kPa), T_{boil} = Normal boiling point temperature (K), T_c = Critical temperature (K), T_{fus} = Normal melting (fusion) point (K).

5.2 Benchmarking and model selection

To establish rigorous baselines and test whether nonlinear learning provides added value beyond a linear Prodeg-to-property mapping, we benchmarked a suite of regression models for predicting physicochemical properties from four degree-based Prodeg descriptors ($Z_{-1/2}$, IP, $\overline{IP}$, MPI). The core endpoint set comprised nine properties: {$McVol$, V_c, $\Delta_f G^\circ$, $\Delta_f H^\circ_{\text{gas}}$, $\log_{10} WS$, P_c, T_{boil}, T_c, T_{fus}}. Three additional compiled properties ($\log P_{\text{oct/wat}}$, $\Delta_{vap} H^\circ$, and $\Delta_{fus} H^\circ$) were excluded from the core benchmark because degree-descriptor-only models produced weak and/or unstable predictability; they are analyzed separately (Suppl. Table 4).

External validation split and leakage control. The full dataset ($n = 90$) was partitioned into a training set ($n = 72$) and an independent test set ($n = 18$) using the Kennard-Stone algorithm in the four-dimensional Prodeg descriptor space (80:20 split). Kennard-Stone selection used *predictors only* (no endpoint information). Standardization for distance computation was applied only in descriptor space. All subsequent preprocessing (including scaling), cross-validation, and hyperparameter tuning were performed *exclusively within the training set.* After training-only model selection, each final configuration was refit on the full training set and evaluated *once* on the fixed external test set. This workflow follows established QSAR/QSPR validation guidance and is designed to prevent test leakage. [64–66]

Candidate models and endpoint-wise fitting strategy. We benchmarked a naive mean predictor (DummyMean), linear/latent-linear baselines (OLS, PLS regression, ElasticNet, and Ridge_perTargetAlpha), and nonlinear learners spanning instance-based, kernel, tree-ensemble, boosted-tree, and neural-network families (Table 6). Because several algorithms are intrinsically single-output and because optimal regularization/complexity may differ across endpoints, we adopted a *one-model-per-endpoint* protocol: nine independent fits under the same fixed split, enabling fair property-wise comparisons.

Preprocessing within cross-validation. For scale-sensitive methods (PLS, ElasticNet/Ridge, SVR, KNN, and MLP), standardization of the four Prodeg descriptors was learned *within each CV fold* (scaler fit on the fold-specific training portion and applied to its validation portion), preventing information leakage. For the final refit, preprocessing parameters were estimated from the full training set and then applied to the external test set.

Hyperparameter tuning and selection criterion. All candidate models were tuned by 5-fold cross-validation on the training set only. Because endpoints have heterogeneous units and variances, CV performance for endpoint j was summarized by the unit-normalized RMSE

$$\text{NRMSE}_j = \frac{\text{RMSE}_j}{s_{j,\text{train}}}, \tag{8}$$

where $s_{j,\text{train}}$ is the standard deviation of endpoint j computed on the *training portion* of that fold. Hyperparameters were selected *per endpoint* by minimizing the corresponding CV NRMSE under identical fold splits. To compare algorithms across endpoints, we report the mean of endpoint-wise best CV NRMSE values (averaged across folds), so that no single high-variance endpoint dominates comparisons (Suppl. Table 2).

Final refit and external reporting. After selecting hyperparameters using training-set CV only, each model was refit on the full training set and evaluated *once* on the fixed Kennard-Stone external test set. External predictivity is summarized primarily by R^2 and also by RMSE and MAE. For compact comparison across algorithms, we report the mean external-test R^2 averaged over the nine endpoints (Table 8), while retaining endpoint-wise metrics for interpretability and diagnosis (Suppl. Table 1). The full benchmark summary (including best hyperparameters) is provided in Suppl. Table 5.

To avoid overstating correlation-based evidence, we interpret model quality through a combined validation profile rather than through training correlations alone. This profile includes external-test R^2, RMSE, and MAE; training-only cross-validation; bootstrap confidence intervals on the fixed test set; Y-randomization; residual diagnostics; and leverage-based applicability-domain analysis. Because degree-based graph descriptors often share size and branching information,

ARTICLE IN PRESS

Table 5. Four Prodeg indices and the nine core physicochemical endpoints used for QSPR modeling (all $n = 90$ compounds). Values are rounded to two decimals.

Formula	$Z_{-1/2}$	IP	$\overline{IP}$	MPI	$McVol$	V_c	$\Delta_f G^\circ$	$\Delta_f H^\circ_{\text{gas}}$	$\log_{10} WS$	P_c	T_{boil}	T_c	T_{fus}
C9H8O4	9.84	18.00	100.14	38.93	128.79	0.48	-371.98	-513.64	-1.70	4082.92	659.32	868.49	322.40
C10H10O4	10.55	19.41	117.76	41.76	142.88	0.54	-363.56	-534.28	-2.26	3624.61	682.20	888.42	424.31
C9H10O3	8.97	16.68	82.04	35.56	127.22	0.47	-243.06	-401.06	-2.07	3773.04	605.45	807.67	363.11
C10H10O4	10.55	19.41	117.76	41.76	142.88	0.54	-331.74	-514.27	-2.03	3191.93	612.44	831.50	385.72
C9H7BrO4	10.72	19.32	119.98	42.29	146.29	0.54	-367.29	-498.78	-3.01	4379.97	730.46	951.00	485.36
C18H14O7	18.40	35.41	406.11	78.68	240.85	0.91	-524.44	-811.71	-4.31	2239.76	957.30	1196.42	636.91
C9H8O5	10.55	19.41	117.76	41.76	134.66	0.50	-476.98	-645.86	-1.91	3995.65	681.74	888.91	435.27
C10H9ClO4	11.42	20.73	139.18	45.12	155.12	0.58	-377.93	-555.30	-2.53	3501.28	719.19	932.38	439.23
C8H8O3	8.27	15.27	67.41	32.73	113.13	0.42	-251.48	-433.80	-1.56	4260.71	582.57	787.82	372.23
C10H10O5	11.42	20.73	139.18	45.12	148.75	0.55	-478.19	-677.97	-1.97	3488.88	709.60	914.96	459.06
C9H8O5	10.55	19.41	117.76	41.76	134.66	0.50	-508.80	-665.87	-1.17	4608.87	751.50	950.28	473.86
C10H7F3O4	13.05	23.41	185.42	51.76	148.19	0.58	-913.33	-1111.35	-2.69	2811.36	607.02	808.77	389.91
C10H7Cl3O5	13.76	24.83	209.08	54.58	185.47	0.69	-501.51	-722.47	-3.89	3191.93	813.68	1037.46	538.72
C9H8O4	9.84	18.00	100.14	38.93	128.79	0.48	-371.98	-513.64	-1.85	4082.92	659.32	868.49	400.00
C9H8O4	9.84	18.00	100.14	38.93	128.79	0.48	-371.98	-513.64	-1.85	4082.92	659.32	868.49	413.04
C14H10O5	13.83	27.12	221.78	60.63	181.35	0.62	-372.09	-557.62	-3.17	4082.92	881.02	1117.60	607.53
C11H12O5	12.13	22.15	159.80	47.95	162.84	0.61	-437.95	-678.60	-2.15	2775.92	662.72	876.45	431.74
C13H14O6	14.25	26.39	230.12	56.44	192.59	0.73	-540.40	-820.99	-2.15	2470.27	757.37	970.44	491.69
C10H12O3	9.68	18.10	98.09	38.39	141.31	0.53	-202.82	-401.69	-2.25	2976.29	558.57	770.06	335.79
C14H8O4	12.54	26.53	186.56	61.46	164.62	0.61	-88.50	-334.59	-3.89	3501.28	788.26	1070.21	533.66
C21H12O6	18.80	39.80	449.07	92.19	246.93	0.91	-159.69	-536.03	-5.98	2576.72	1083.39	1383.53	739.24
C11H7F5O4	15.55	27.41	268.08	61.76	165.82	0.66	-1291.69	-1532.96	-3.42	2372.59	625.21	817.70	404.78
C10H10O5	11.26	20.83	136.80	44.58	148.75	0.55	-436.74	-646.49	-2.09	3131.49	634.86	851.27	407.95
C13H8F6O6	19.25	34.39	427.63	76.44	203.21	0.81	-1703.58	-2015.15	-3.47	1985.89	746.53	936.07	500.07
C11H14O3	10.55	19.41	117.76	41.76	155.40	0.58	-196.84	-427.61	-2.78	2709.85	581.01	793.47	332.06
C9H7ClO3	9.84	18.00	100.14	38.93	135.16	0.51	-247.09	-377.15	-2.67	3484.76	604.57	834.38	382.14
C14H12O3	12.09	24.48	168.92	53.90	173.91	0.65	-56.73	-247.72	-3.68	2811.36	676.77	916.60	407.29
C11H14O3	10.39	19.51	115.55	41.22	155.40	0.59	-194.40	-422.33	-2.67	2687.42	581.45	789.55	347.06
C9H5F3O4	12.34	22.00	163.17	48.93	134.10	0.52	-953.57	-1110.72	-2.51	3538.87	653.90	848.20	417.23
C9H10O4	9.84	18.00	100.14	38.93	133.09	0.49	-357.69	-544.75	-1.77	3628.97	632.85	834.15	382.50
C17H15NO5	16.82	32.68	337.35	72.49	229.30	0.86	-209.55	-492.80	-4.07	2379.54	908.30	1145.45	606.14
C15H14O4	13.66	27.22	218.75	60.09	193.87	0.72	-162.94	-412.05	-3.80	2465.36	727.05	959.54	453.31
C11H9F3O5	14.63	26.15	237.18	57.95	168.15	0.65	-1019.54	-1275.68	-2.81	2465.36	657.30	855.60	435.93
C12H7F7O4	18.05	31.41	365.73	71.76	183.45	0.74	-1670.05	-1954.57	-4.16	2029.06	643.40	828.09	419.65
C18H14O3	14.66	30.78	262.35	69.95	210.81	0.79	73.97	-150.68	-5.48	2402.92	792.25	1040.33	497.59
C9H9NO3	9.84	18.00	100.14	38.93	132.90	0.49	-168.71	-327.62	-1.96	3906.25	639.67	872.95	435.48
C8H7ClO3	9.14	16.59	83.93	36.10	125.37	0.47	-273.04	-407.63	-2.34	3985.56	624.98	835.06	394.28
C14H11ClO3	12.96	25.80	194.48	57.27	186.15	0.69	-78.29	-274.93	-4.36	2662.52	719.18	962.28	449.73
C15H12O5	14.54	28.53	247.63	63.46	195.44	0.67	-331.85	-558.25	-3.42	3191.93	834.14	1080.69	580.21
C9H10O4	9.84	18.00	100.14	38.93	133.09	0.49	-357.69	-578.60	-1.77	3628.97	632.85	834.15	397.86
C15H12O3	12.96	25.80	194.48	57.27	183.70	0.69	-72.23	-248.72	-3.74	2764.26	731.10	975.09	446.26
C7H6O3	7.56	13.85	54.19	29.90	99.04	0.31	-299.89	-492.40	-1.89	6349.11	612.91	833.23	431.35
C11H14O3	10.39	19.51	115.55	41.22	155.40	0.59	-194.40	-422.33	-2.67	2687.42	581.45	789.55	347.06
C15H11NO3	13.66	27.22	218.75	60.09	189.38	0.73	75.24	-114.95	-4.03	2445.89	806.71	1052.53	496.07
C13H18O3	12.13	22.15	159.80	47.95	183.58	0.69	-182.44	-474.17	-3.73	2256.81	626.33	836.17	339.60
C17H18O4	15.24	29.95	274.90	66.29	222.05	0.83	-148.54	-458.61	-4.75	2069.88	772.37	1000.99	460.85
C9H10O4	9.84	18.00	100.14	38.93	133.09	0.49	-357.69	-544.75	-2.14	3628.97	632.85	834.15	397.86
C9H8O4	9.84	18.00	100.14	38.93	128.79	0.48	-371.98	-513.64	-1.85	4082.92	659.32	868.49	413.04
C14H10Cl2O3	13.83	27.12	221.78	60.63	198.39	0.74	-99.85	-302.14	-5.05	2525.19	761.59	1007.60	492.17
C13H10O3	11.38	23.07	147.64	51.07	159.82	0.54	-105.14	-260.70	-3.15	3862.67	707.11	962.79	314.75
C12H9F5O5	17.13	30.15	329.57	67.95	185.78	0.73	-1397.90	-1697.29	-3.54	2102.27	675.49	865.86	450.80
C10H10O4	10.55	19.41	117.76	41.76	142.88	0.54	-331.74	-514.27	-2.03	3191.93	612.44	831.50	385.72
C14H11ClO3	12.96	25.80	194.48	57.27	186.15	0.69	-78.29	-274.93	-4.36	2662.52	719.18	962.28	449.73
C10H12O3	9.68	18.10	98.09	38.39	141.31	0.53	-202.82	-401.69	-2.25	2976.29	534.65	770.06	335.79
C10H5F5O4	14.84	26.00	241.21	58.93	151.73	0.60	-1331.93	-1532.33	-3.24	2928.17	672.09	858.93	432.10
C14H10F2O3	13.83	27.12	221.78	60.63	177.45	0.68	-465.61	-662.88	-4.34	2485.07	685.27	905.92	433.51
C15H11ClO3	13.83	27.12	221.78	60.63	195.94	0.74	-93.79	-275.93	-4.43	2619.09	773.51	1020.33	488.70
C14H12O4	12.96	25.80	194.48	57.27	179.78	0.61	-211.35	-425.03	-3.23	3314.37	757.39	1004.09	519.01
C10H10O4	10.55	19.41	117.76	41.76	142.88	0.54	-331.74	-514.27	-2.03	3191.93	612.44	831.50	385.72
C15H8F10O6	24.25	42.39	685.13	96.44	238.47	0.98	-2460.30	-2858.37	-4.94	1504.65	782.91	966.28	529.81
C15H12O3	12.96	25.80	194.48	57.27	183.70	0.69	-72.23	-248.72	-4.22	2764.26	731.10	975.09	446.26
C13H9ClO3	12.25	24.39	171.62	54.44	172.06	0.59	-126.70	-287.91	-3.79	3624.61	749.52	1007.84	515.43
C14H11NO5	14.54	28.53	247.63	63.46	191.33	0.73	-30.81	-269.95	-4.33	2802.44	833.59	1091.42	563.42
C12H16O3	11.26	20.83	136.80	44.58	169.49	0.64	-188.42	-448.25	-3.20	2458.04	603.89	812.91	343.33
C11H5F7O4	17.34	30.00	334.24	68.93	169.36	0.69	-1710.29	-1953.94	-3.97	2462.92	690.28	871.19	446.97
C15H14O3	12.96	25.80	194.48	57.27	188.00	0.70	-57.94	-279.83	-4.15	2507.52	704.63	940.44	431.08
C12H16O3	11.26	20.83	136.80	44.58	169.49	0.64	-188.42	-448.25	-2.85	2458.04	603.89	812.91	343.33
C16H12O6	15.69	32.00	297.49	73.85	204.54	0.76	-300.92	-663.25	-4.13	2640.67	888.82	1149.83	625.70
C9H10O3	8.97	16.68	82.04	35.56	127.22	0.47	-211.24	-381.05	-1.83	3314.37	521.20	750.72	324.52
C11H10O5	12.13	22.15	159.80	47.95	158.54	0.60	-452.24	-647.49	-2.22	3076.16	689.19	910.86	446.92
C10H10O4	10.55	19.41	117.76	41.76	142.88	0.55	-311.97	-498.74	-2.15	3177.55	612.21	827.41	390.31
C20H18O7	19.81	38.24	476.85	84.34	269.03	1.02	-507.60	-852.99	-4.16	1882.17	1003.06	1240.69	659.45
C18H14O7	18.07	35.61	398.09	77.60	240.85	0.93	-465.64	-757.71	-3.32	2287.14	946.88	1179.94	621.05
C10H7F3O4	13.05	23.41	185.42	51.76	148.19	0.58	-913.33	-1111.35	-2.69	2811.36	607.02	808.77	389.91
C21H18O4	17.48	36.43	383.20	81.26	254.65	0.95	-0.01	-299.36	-5.91	1959.60	891.01	1137.48	547.35
C13H18O3	11.80	22.34	154.71	46.87	183.58	0.70	-177.56	-463.61	-3.51	2222.89	627.21	829.16	369.60
C14H10BrFO3	13.83	27.12	221.78	60.63	193.18	0.73	-256.48	-440.44	-5.17	2793.56	752.16	992.93	492.72
C12H12O5	13.00	23.46	184.53	51.32	172.63	0.65	-453.45	-679.60	-2.76	2729.71	717.05	935.93	470.71
C10H12O4	10.39	19.51	115.55	41.22	147.18	0.55	-339.64	-553.92	-1.52	3299.15	650.75	848.60	396.61
C15H14O3	12.96	25.80	194.48	57.27	188.00	0.70	-57.94	-279.83	-4.15	2507.52	704.63	940.44	431.08
C18H12O6	17.36	34.19	365.17	74.77	230.68	0.89	-280.42	-508.27	-4.08	2445.89	928.62	1168.44	593.74
C13H9F7O5	19.63	34.15	436.96	77.95	203.41	0.81	-1776.26	-2118.90	-4.28	1813.86	693.68	877.67	465.67
C19H16O6	18.07	35.61	398.09	77.60	249.07	0.97	-352.22	-646.13	-4.65	2123.64	947.34	1179.79	610.09
C16H14O5	15.24	29.95	274.90	66.29	209.53	0.78	-283.44	-545.27	-3.92	2386.52	803.80	1037.51	514.51
C13H16O5	13.38	25.07	202.40	53.07	191.02	0.72	-411.48	-708.41	-2.37	2322.54	703.50	910.34	441.76
C13H10O3	11.38	23.07	147.64	51.07	159.82	0.59	-96.97	-247.09	-3.17	3538.87	723.65	951.31	434.61
C9H10O4	9.84	18.00	100.14	38.93	133.09	0.49	-357.69	-544.75	-1.77	3628.97	632.85	834.15	397.86
C14H12O3	12.25	24.39	171.62	54.44	173.91	0.59	-106.35	-292.81	-3.58	3380.21	734.97	985.71	496.78
C11H14O4	11.10	20.92	134.42	44.04	161.27	0.60	-299.40	-554.55	-1.76	2640.67	603.87	809.83	369.29
C11H14O3	10.55	19.41	117.76	41.76	155.40	0.58	-196.84	-427.61	-2.78	2709.85	581.01	793.47	332.06

ARTICLE IN PRESS

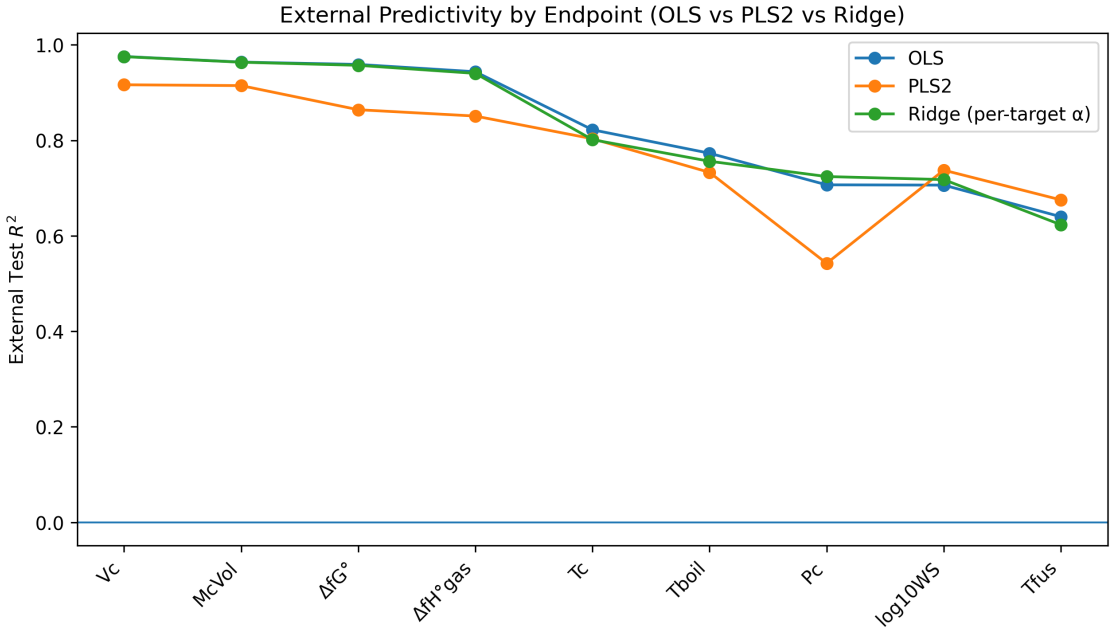

Figure 1. Model comparison via external-test R^2 (mean across nine endpoints) under the fixed Kennard–Stone split. The dashed reference line corresponds to zero external predictivity; higher bars indicate stronger average external generalization across the nine benchmarked endpoints.

individual standardized coefficients are interpreted as associative indicators rather than causal effects, and endpoint-level accuracy is emphasized over isolated descriptor-response correlations.

Benchmark outcome and model ranking. Across the nine-endpoint task, linear/latent-linear models provided the strongest overall generalization. OLS and PLS2 achieved the top mean external-test performance (Table 8), with Ridge/α close behind; see Figure 1. In contrast, nonlinear learners did not improve *average* external predictivity. Overall, with only four degree-based Prodeg descriptors, the dominant structure-property signal for the selected endpoints is captured by a near-linear mapping; additional model flexibility does not translate into reproducible gains on the external test set. These results should be read as supportive evidence for descriptor relevance within this dataset, not as definitive validation of universal predictive superiority.

Final reporting set. Given the near-tie among OLS, PLS2, and Ridge/α, we retained these three models as the final reporting set. OLS provides a maximally transparent reference; PLS2 provides a latent-linear cross-check; and Ridge/α provides endpoint-specific shrinkage that can improve stability under mild noise or collinearity. Endpoint-wise external R^2 values for these three models are summarized in Table 7, while complete endpoint-wise metrics for all benchmarked algorithms are reported in Suppl. Table 1.

Benchmark outcome and model ranking. Table 8 summarizes the compact benchmark under the fixed Kennard-Stone external test set, with all tuning performed using training-set cross-validation only (no test leakage). Across the nine-endpoint task, linear/latent-linear models provided the strongest overall generalization: OLS and PLS2 achieved identical mean external-test performance (mean $R^2 = 0.832$), followed closely by Ridge/α (mean $R^2 = 0.829$). In contrast, nonlinear learners did not improve average external predictivity (best nonlinear mean $R^2 = 0.643$ for GBR; MLP mean $R^2 = 0.632$). Complete per-endpoint metrics for all benchmarked models are reported in Suppl. Table 1.

Final reporting set. Given the near-tie among OLS, PLS2, and Ridge/α , we retained these three models as the final reporting set. OLS provides a transparent reference mapping, PLS2 provides a latent-linear cross-check, and Ridge/α provides endpoint-specific shrinkage that can improve stability under mild noise or collinearity.

5.3 External validation, robustness, applicability domain, and interpretability

External predictivity and pointwise agreement. For the final reporting set (OLS, PLS2, and Ridge/α), external predictivity was assessed *only* on the fixed Kennard-Stone test set (held out throughout tuning and model selection). Figure 2 shows predicted versus observed values on the external test set for the reference model (OLS; “global_best” in our

ARTICLE IN PRESS

Table 6. Regression models benchmarked for QSPR prediction from the four Prodeg indices ($Z_{-1/2}$, IP, $\overline{IP}$, MPI), grouped by model family. One independent model was fit per endpoint (nine targets) to enable a consistent comparison.

Model family	**Benchmarked algorithms (abbreviation in this work)**
Naïve baseline	Mean predictor (DummyMean)
Linear / latent linear baselines	Ordinary least squares multiple linear regression (OLS); Partial least squares regression (PLS2 label used in tables; number of latent components tuned); Elastic net regression (ElasticNet); Ridge regression with per-endpoint regularization strength tuned by cross-validation (Ridge_perTargetAlpha)
Instance-based learning	k-nearest neighbors regression (KNN)
Kernel methods	Support vector regression with RBF kernel (SVR_RBF); ν-support vector regression with RBF kernel (NuSVR_RBF); Linear support vector regression (LinearSVR)
Tree ensembles (bagging)	Random forest regressor (RandomForest); Extremely randomized trees (ExtraTrees); Bagging of decision trees (BaggingTree)
Boosted trees	Gradient boosting regressor (GBR); Histogram-based gradient boosting regressor (HistGB)
Neural networks	Multilayer perceptron regressor (MLP)

Table 7. External test-set performance for nine physical properties predicted from the four Prodeg indices ($Z_{-1/2}$, IP, $\overline{IP}$, MPI) using (i) OLS, (ii) PLS2 (per-target tuned number of components k), and (iii) Ridge_perTargetAlpha (per-target tuned α). Entries report R^2 on the external test set; for OLS and Ridge, parentheses give (RMSE/MAE). The best R^2 per property is highlighted in bold.

Property	OLS R^2 (RMSE/MAE)	PLS2 R^2 (k)	Ridge R^2 (RMSE/MAE, α)
$McVol$	**0.964** (8.682/7.265)	**0.964** $k = 4$	0.964 (8.714/7.121) $\alpha = 0.0106$
V_c	**0.975** (0.0286/0.0231)	**0.975** $k = 4$	0.975 (0.0288/0.0216) $\alpha = 0.0106$
$\Delta_f G^\circ$	**0.959** (141.4/113.6)	**0.959** $k = 4$	0.957 (144.6/118.9) $\alpha = 0.006649$
$\Delta_f H^\circ_{\mathrm{gas}}$	**0.944** (170.9/149.6)	**0.944** $k = 4$	0.940 (176.0/154.7) $\alpha = 0.008396$
$\log_{10} WS$	0.706 (0.676/0.505)	0.706 $k = 4$	**0.718** (0.6629/0.5037) $\alpha = 0.03401$
P_c	0.707 (588.7/393.7)	0.707 $k = 3$	**0.724** (570.1/390.0) $\alpha = 0.1505$
T_{boil}	**0.773** (65.35/52.85)	0.733 $k = 3$	0.756 (67.71/54.68) $\alpha = 0.03401$
T_c	**0.822** (68.23/59.78)	0.803 $k = 2$	0.801 (69.17/60.50) $\alpha = 0.01266$
T_{fus}	0.640 (63.57/53.68)	**0.675** $k = 4$	0.623 (64.99/54.82) $\alpha = 0.003429$

Table 8. Compact benchmark comparison of representative regressors for predicting the nine endpoints from the four Prodeg indices. CV scores are computed on the training set only; external test performance is summarized by the mean R^2 across endpoints.

Model	**Train CV mean NRMSE↓**	**External test mean R^2 ↑**
DummyMean	1.013	-0.263
OLS	0.476	**0.832**
PLS2	0.476	**0.832**
Ridge_per_Target_Alpha	0.464	**0.829**
GBR (best boosted tree)	0.688	0.643
MLP (neural network)	0.619	0.632

Note: Ridge_perTargetAlpha is tuned separately per endpoint; the reported CV mean NRMSE is the average (across endpoints) of each endpoint's best-CV configuration under the same fold splits.

pipeline), displayed separately for the nine benchmarked endpoints. Agreement is strongest for several targets and more moderate for others, with T_{fus} remaining the most challenging endpoint, consistent with weaker signal-to-noise under a

ARTICLE IN PRESS

degree-descriptor-only representation for this aromatic-carboxylate series.

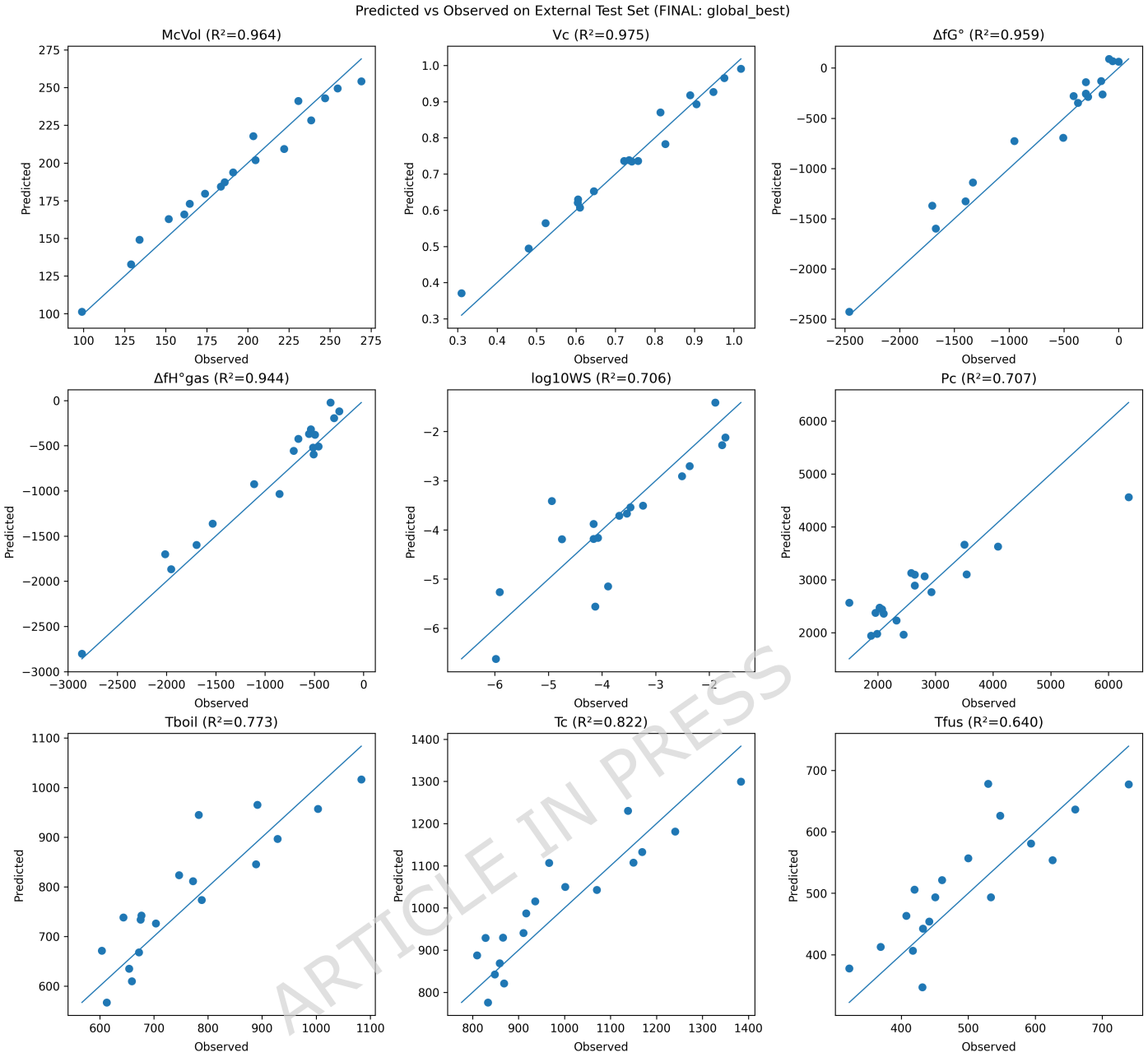

Figure 2. Predicted versus observed values on the external Kennard-Stone test set for the reference model (OLS; global_best), shown for the nine benchmarked endpoints. The diagonal indicates perfect agreement.

Residual diagnostics and error structure. To check for systematic bias and gross misspecification, we examined residuals on the external test set (Figure 3), plotting $(y_{\text{obs}} - y_{\text{pred}})$ against y_{pred} for each endpoint. Residuals are generally centered near zero without a consistent trend across the modeled range, supporting the benchmark finding that the dominant structure-property signal captured by these four descriptors is close to linear. Dispersion is endpoint-dependent and amplified by the modest external set size ($n_{\text{test}} = 18$).

Uncertainty quantification via bootstrap confidence intervals. Because the external test set is limited ($n_{\text{test}} = 18$), we quantified uncertainty in external performance using a nonparametric bootstrap over the *test compounds* ($B = 2000$ resamples with replacement), recomputing test metrics for each resample (model fixed; no refitting). Bootstrap 95% confidence intervals for external-test R^2, RMSE, and MAE are reported in Supplementary Table S3, defined in the Supplementary Information as "Bootstrap confidence intervals for external-test metrics under the fixed Kennard–Stone split".

Y-randomization as a chance-correlation control. To rule out spurious correlations, we performed Y-randomization by permuting the training responses and rerunning the same training-only workflow under the identical fixed Kennard-

ARTICLE IN PRESS

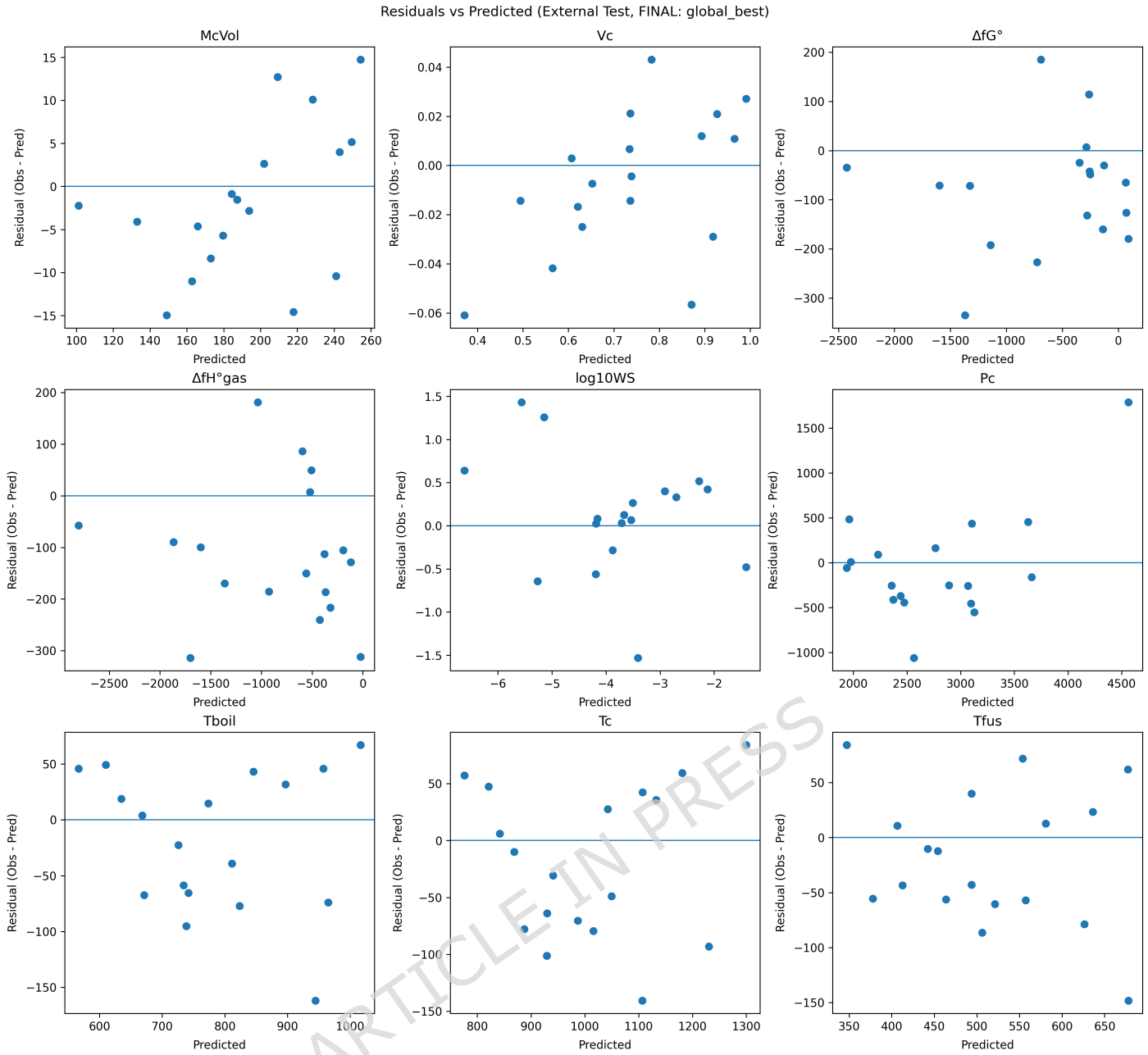

Figure 3. External-test residual diagnostics for the nine benchmarked endpoints under the reference model (OLS). Panels show residuals ($y_{\text{obs}} - y_{\text{pred}}$) versus y_{pred}, highlighting endpoint-dependent dispersion and checking whether prediction errors show systematic trends across the modeled range.

Stone split ($N = 200$ permutations; predictors unchanged). The null distribution of mean external-test R^2 (averaged across the nine endpoints) is centered near and below zero, and none of the randomized runs matched the observed performance for the true labels (Figure 4). This supports that the reported predictivity is not attributable to chance alignment between descriptors and responses.

Interpretability via standardized coefficients and descriptor influence. Given the near-tie among the final linear models, we interpret the mapping primarily through standardized OLS coefficients. Predictors were z-scored using parameters learned on the training set, and coefficients are reported in standardized form to enable comparison across descriptors and endpoints. The standardized coefficient matrix across endpoints is shown in Figure 5, and a compact global summary (mean absolute standardized coefficient across endpoints) is shown in Figure 6. Overall, IP contributes most strongly on average, followed by MPI and $Z_{-1/2}$, while $\overline{IP}$ contributes more weakly across most targets (associative, not causal).

Applicability domain (Williams plot). Model reliability with respect to descriptor-space extrapolation was assessed using a Williams plot for a representative endpoint (V_c) under OLS (Figure 7). Let X_{train} denote the standardized training design matrix and x the standardized descriptor vector for a given compound. Leverage was computed as

ARTICLE IN PRESS

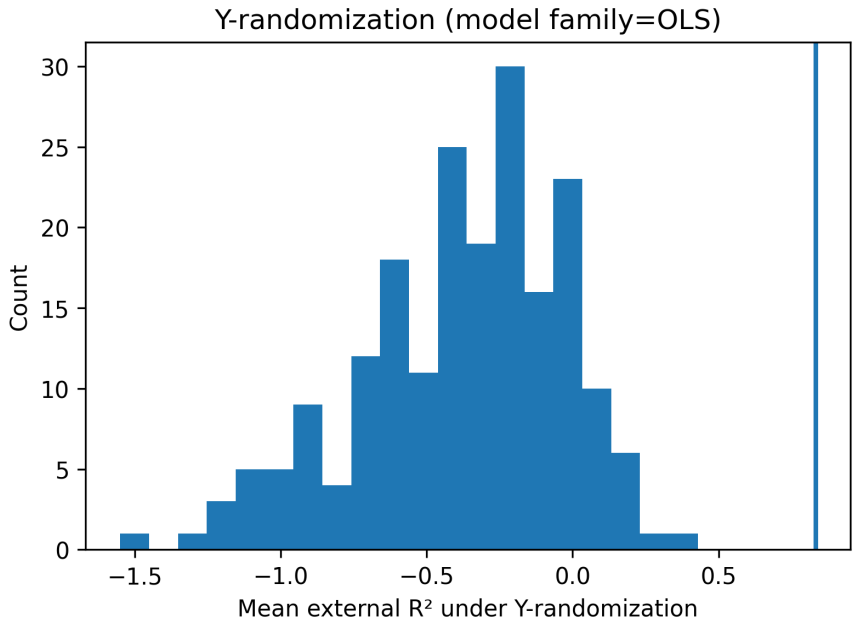

Figure 4. Y-randomization test for the OLS model family. Histogram shows the distribution of mean external-test R^2 across $N = 200$ permutations of the training responses under the fixed Kennard-Stone split; the vertical line indicates the observed mean external-test R^2 for the true labels.

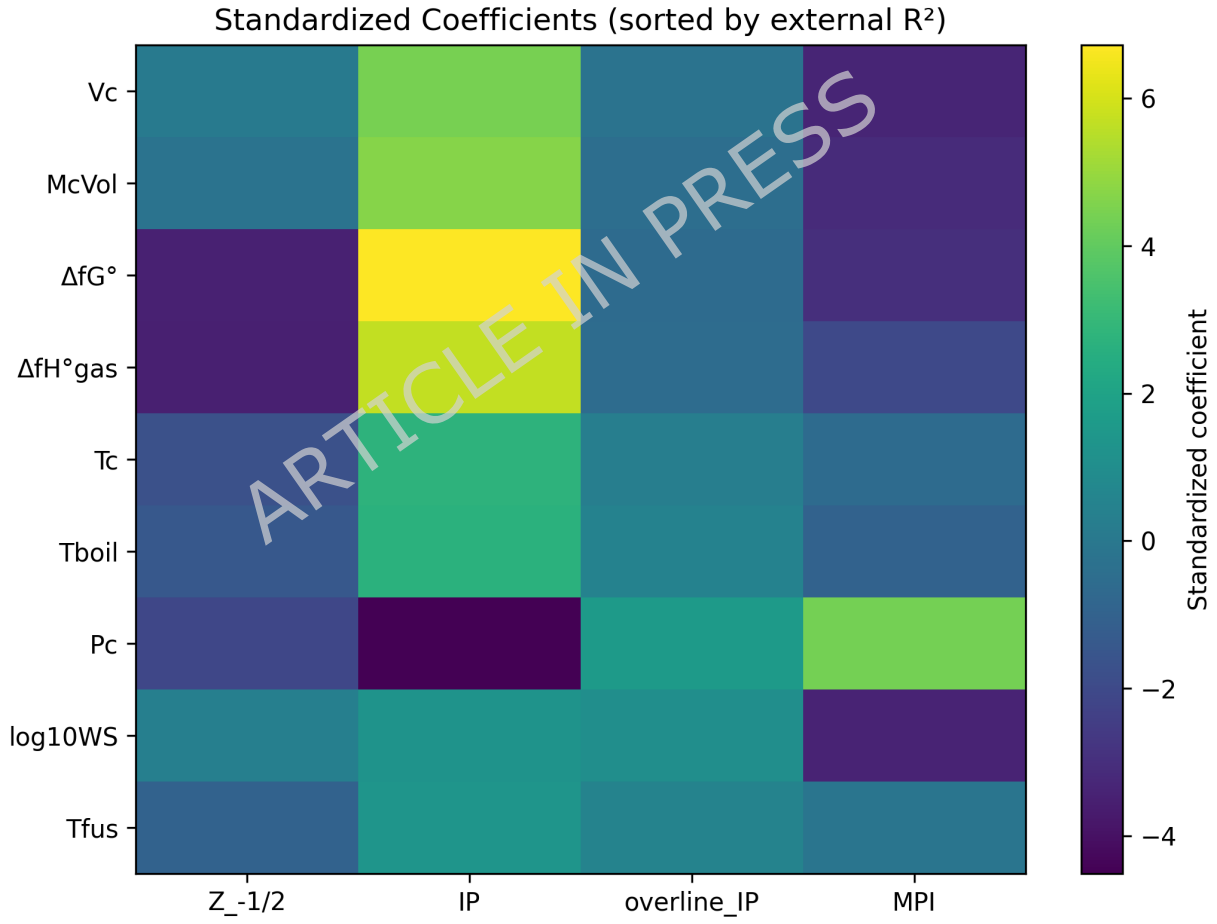

Figure 5. Standardized OLS coefficients across endpoints for the four Prodeg descriptors ($Z_{-1/2}$, IP, $\overline{IP}$, MPI). Coefficients are computed after standardizing predictors, enabling comparison of effect sizes across descriptors within each endpoint.

$h = x^\top (X_{\text{train}}^\top X_{\text{train}})^{-1} x$. The warning leverage threshold was set to $h^* = 3(p+1)/n_{\text{train}}$ with $p = 4$ and $n_{\text{train}} = 72$, giving $h^* \approx 0.208$. Standardized residual reference limits of ± 3 were used to flag potential response outliers. Some external test compounds may approach or exceed h^* due to the deliberate coverage of descriptor-space extremes by the Kennard-Stone split; such cases should be interpreted as extrapolative in descriptor space and treated with increased caution.

ARTICLE IN PRESS

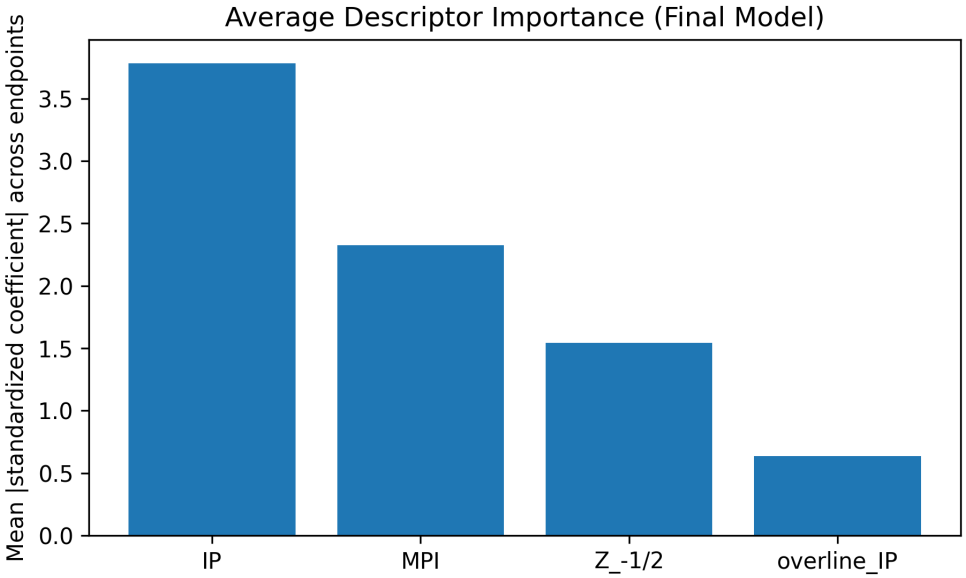

Figure 6. Average descriptor importance for the final model, computed as the mean absolute standardized coefficient across the nine endpoints.

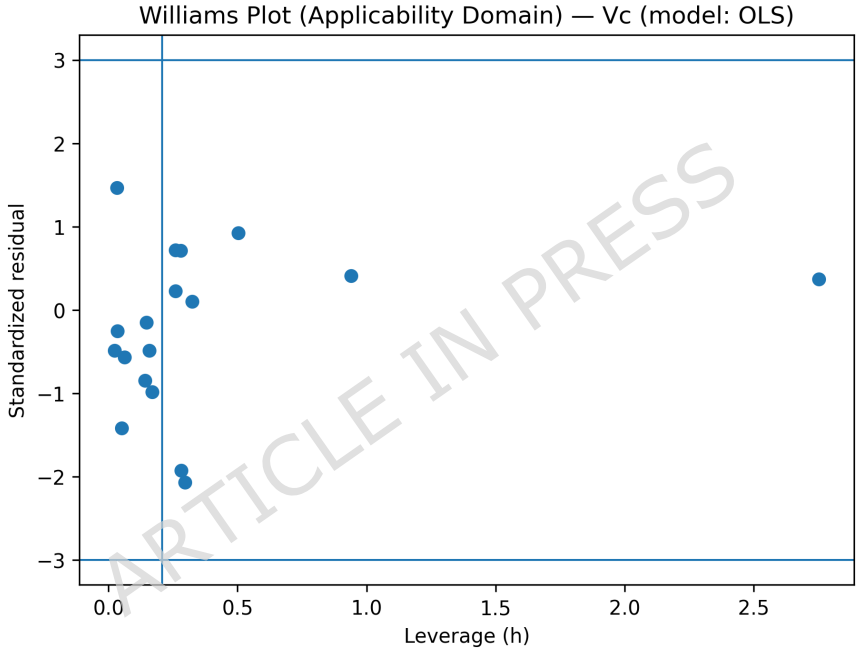

Figure 7. Williams plot (applicability domain) for a representative endpoint (V_c) under OLS. The vertical line indicates the warning leverage threshold h^* and horizontal lines show standardized residual reference limits ± 3; points beyond these limits require cautious interpretation as potential extrapolative or response-outlier cases.

Conclusion

Mathematical contribution. We introduced the *Inverse Prodeg* index $IP(G)$ and its *coindex* $\overline{IP}(G)$ as degree-based invariants with applications in extremal graph theory and cheminformatics. We established sharp bounds in terms of order, size, and degree extrema, including equality characterizations for regular and extremal graph classes. We also obtained Nordhaus–Gaddum-type sum and product inequalities for $IP(G)$ and $\overline{IP}(G)$, exact values for several classical families (including paths, cycles, stars, complete and multipartite graphs, wheels, ladders, hypercubes, and friendship graphs), and estimates under common graph operations such as Cartesian, strong, disjunction, symmetric difference, join, and tensor product. The corrected tree result shows that, among trees of fixed order, the star minimizes IP and the path maximizes it. The index is computable in linear time in the number of edges, which facilitates large-scale applications.

Empirical QSPR findings. To assess empirical utility in QSPR, we constructed a curated dataset of $n = 90$ aromatic-carboxylate compounds and compiled physicochemical endpoints from Cheméo. Using four degree-based Prodeg descriptors $\{Z_{-1/2}, IP, \overline{IP}, MPI\}$ computed from hydrogen-suppressed molecular graphs, we performed an external-validation benchmark under a fixed Kennard–Stone 80:20 split with training-only preprocessing, cross-validation, and hyperparam-

ARTICLE IN PRESS

eter tuning. The dataset audit corrected the endpoint label for the aspirin thermodynamic value; the relevant column is treated as standard enthalpy of fusion rather than gas-phase enthalpy of formation.

Across the nine-endpoint benchmark, linear/latent-linear mappings provided the most reproducible external predictivity among the candidate models: OLS and PLS achieved identical mean external-test performance (mean $R^2 = 0.832$), with Ridge regression close behind (mean $R^2 = 0.829$). Nonlinear learners did not improve *average* external-test performance in this restricted Prodeg-only setting. Endpoint-wise, predictability was strongest for size- and thermodynamics-related targets and weakest for T_{fus}, consistent with the sensitivity of melting behavior to effects not fully encoded by degree-only topology. Robustness checks, including bootstrap confidence intervals, Y-randomization, residual diagnostics, and applicability-domain analysis, support that the observed predictivity is not merely a chance correlation, but it remains most reliable within the descriptor-space domain covered by the training set.

Limitations and future work. The empirical results should be interpreted as supportive evidence that $IP(G)$ and related Prodeg descriptors encode useful degree-based information, not as a universal superiority claim over established topological, constitutional, electronic, 3D, or fingerprint-based descriptors. The reported values are also conditional on the curated Cheméo records, the fixed Kennard–Stone split, and the restricted aromatic-carboxylate chemical domain. The dataset is small and chemically coherent, which improves interpretability but limits generalizability. Future research should benchmark $IP(G)$ and $\overline{IP}(G)$ against classical descriptors such as Zagreb, Randić, and Sombor families across larger and chemically diverse datasets, extend the analysis to additional acyclic/tree families and weighted or directed graphs, and integrate these descriptors with chemically enriched representations and interpretable machine-learning tools.

The $IP(G)$ index therefore provides a mathematically tractable and computationally efficient descriptor for research at the interface of discrete mathematics, network science, and cheminformatics.

Data availability

Physicochemical property values for the $n = 90$ compounds were compiled from Cheméo (`https://www.chemeo.com/`) on 07 Jan 2026 (units as reported). The curated dataset used for QSPR modeling includes molecular identifiers (molecular name, InChI, InChIKey, molecular formula, canonical SMILES), four degree-based Prodeg descriptors ($Z_{-1/2}$, IP, $\overline{IP}$, MPI), and the compiled physicochemical endpoints. Record identifiers and additional figures/tables are provided in the Supplementary Materials.

The curated dataset and reproducible code are deposited on Zenodo under the Creative Commons Attribution 4.0 International license (CC BY 4.0) as: *Alsharafi, M., & ZEREN, Y. (2026). Degree-Based Prodeg Indices in Graph Theory and Cheminformatics with Sharp Results and QSPR Modeling: Data and Codes. Zenodo.* DOI: `https://doi.org/10.5281/zenodo.18694027`. This record contains (i) the Excel dataset file `Prodeg_QSPR_dataset_Chemeo_2026-01-07_v1.xlsx`, (ii) the benchmarking notebook `ip-prodeg-index-qspr.ipynb` implementing the full QSPR workflow (Kennard–Stone 80:20 split, training-only preprocessing, 5-fold cross-validation and hyperparameter tuning, final refit, and external-test evaluation reporting R^2, RMSE, and MAE), and (iii) the descriptor-computation notebook `Prodeg_Indices_calculation.ipynb` for computing Prodeg indices from hydrogen-suppressed molecular graphs.

Conflict of interest

There is no conflict of interest.

Funding

This research received no funding.

Authors' Contributions

Mohammed Alsharafi: Conceptualization; Investigation; Data curation; Validation; Methodology; Writing - original draft. **Yusuf Zeren:** Conceptualization; Methodology; Formal analysis; Resources; Writing - review & editing. Both authors approved the final version of the manuscript.

ARTICLE IN PRESS

References

[1] I. Gutman, N. Trinajstić, Graph theory and molecular orbitals. Total π-electron energy of alternant hydrocarbons, *Chem. Phys. Lett.* **17** (1972) 535-538.

[2] I. Gutman, Degree-based topological indices, *Croat. Chem. Acta* **68** (1975) 1-8.

[3] S. Nikolić, G. Kovačević, A. Miličević, N. Trinajstić, The Zagreb indices 30 years after, *Croat. Chem. Acta* **76** (2003) 113-124.

[4] K. Xu, M. Liu, K. C. Das, I. Gutman, B. Furtula, A survey on Zagreb indices, *MATCH Commun. Math. Comput. Chem.* **78** (2017) 17-56.

[5] K. C. Das, Sharp bounds for Zagreb indices, *Chem. Phys. Lett.* **511** (2011) 452-454.

[6] X. Zhu, Q. Huang, H. Deng, New generalizations of the Zagreb indices and their extremal properties, *J. Math. Chem.* **58** (2020) 1095-1114.

[7] A. Gagarin, K. Chellali, O. D. Makinde, Generalized Zagreb indices and their applications, *Discrete Math. Lett.* **8** (2022) 30-38.

[8] J. Li, Y. Shi, I. Gutman, Nordhaus-Gaddum-type results for Zagreb indices, *Discrete Appl. Math.* **284** (2020) 396-402.

[9] M. Alsharafi, A. Alameri, The F-index and coindex of V-Phenylenic nanotubes and nanotorus and their molecular complement graphs, *Nanosyst. Phys. Chem. Math.* **12** (2021) 263–270.

[10] M. Alsharafi, Y. Zeren, A. Alameri, The second hyper-Zagreb index of complement graphs and its applications of some nanostructures, *Asian J. Probab. Stat.* **15** (2021) 54–75. https://doi.org/10.9734/ajpas/2021/v15i430364

[11] A. Alameri, M. Alsharafi, Topological indices types in graphs and their applications, *Mathematics Research Developments* (2021).

[12] R. Todeschini, V. Consonni, *Molecular Descriptors for Chemoinformatics*, 2nd ed., Wiley-VCH, Weinheim, 2009.

[13] M. Randić, On characterization of molecular branching, *J. Am. Chem. Soc.* **97** (1975) 6609-6615.

[14] X. Li, I. Gutman, *Mathematical Aspects of Randić-Type Molecular Structure Descriptors*, Univ. Kragujevac, Kragujevac, 2006.

[15] B. Zhou, N. Trinajstić, On general sum-connectivity index, *J. Math. Chem.* **46** (2009) 283-293.

[16] L. Zhong, The harmonic index for graphs, *Appl. Math. Lett.* **25** (2012) 561-566.

[17] E. Estrada, L. Torres, L. Rodríguez, I. Gutman, An atom-bond connectivity index: Modelling the enthalpy of formation of alkanes, *Indian J. Chem., Sect. A* **37A** (1998) 849-855.

[18] D. Vukičević, B. Furtula, Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges, *J. Math. Chem.* **46** (2009) 1369-1376.

[19] D. Vukičević, Bond additive modeling 1: Adriatic indices, *Croat. Chem. Acta* **83** (2010) 243-260.

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

[21] H. Chen, R. Liu, B. Liu, The inverse sum indeg index of graphs with some given parameters, *Int. J. Algebra Comput.* **28** (2018) 269-285.

[22] V. R. Kulli, I. Gutman, On some mathematical properties of Nirmala index, *Ann. Pure Appl. Math.* **23** (2021) 93-99.

[23] V. R. Kulli, V. Lokesha, K. Nirupadi, Computation of inverse Nirmala indices of certain nanostructures, *Int. J. Math. Combin.* **2021**(2) (2021) 33-40.

[24] M. H. Khalifeh, H. Yousefi-Azari, A. R. Ashrafi, The hyper-Wiener index of graph operations, *Comput. Math. Appl.* **56** (2008) 1402-1407.

[25] M. H. Khalifeh, H. Yousefi-Azari, A. R. Ashrafi, The first and second Zagreb indices of some graph operations, *Discrete Appl. Math.* **157** (2009) 804-811.

[26] A. R. Ashrafi, T. Došlić, A. Hamzeh, The Zagreb coindices of graph operations, *Discrete Appl. Math.* **158** (2010) 1571-1578.

[27] K. C. Das, A. Yurttas, M. Togan, A. S. Cevik, I. N. Cangul, The multiplicative Zagreb indices of graph operations, *J. Inequal. Appl.* **2013** (2013) 90.

[28] M. Azari, A. Iranmanesh, Computing the eccentric-distance sum for graph operations, *Discrete Appl. Math.* **161** (2013) 2827-2840.

ARTICLE IN PRESS

ARTICLE IN PRESS

[29] M. Azari, A. Iranmanesh, Chemical graphs constructed from rooted product and their Zagreb indices, *MATCH Commun. Math. Comput. Chem.* **70** (2013) 901-919.

[30] M. Azari, Sharp lower bounds on the Narumi-Katayama index of graph operations, *Appl. Math. Comput.* **239** (2014) 409-421.

[31] M. Azari, A. Iranmanesh, Some inequalities for the multiplicative sum Zagreb index of graph operations, *J. Math. Inequal.* **9** (2015) 727-738.

[32] N. De, A. Pal, S. M. A. Nayeem, The irregularity of some composite graphs, *Int. J. Appl. Comput. Math.* **2** (2016) 371-378.

[33] N. De, A. Pal, S. M. A. Nayeem, On some bounds and exact formulae for connective eccentric indices of graphs under some graph operations, *Int. J. Combin.* **2014** (2014) 579257.

[34] M. Tavakoli, F. Rahbarnia, A. R. Ashrafi, Some new results on irregularity of graphs, *J. Appl. Math. Inform.* **32** (2014) 675-685.

[35] B. Eskender, E. Vumar, Eccentric connectivity index and eccentric distance sum of some graph operations, *Trans. Combin.* **2** (2013) 103-111.

[36] M. Veylaki, M. J. Nikmehr, H. A. Tavallaee, The third and hyper-Zagreb coindices of some graph operations, *J. Appl. Math. Comput.* **52** (2016) 265-279.

[37] L. Barrière, C. Dalfó, M. A. Fiol, M. Mitjana, The generalized hierarchical product of graphs, *Discrete Math.* **309** (2009) 3871-3881.

[38] M. Eliasi, A. Iranmanesh, Hosoya polynomial of hierarchical product of graphs, *MATCH Commun. Math. Comput. Chem.* **69** (2013) 111-119.

[39] M. Alsharafi, M. Shubatah, A. Alameri, The hyper-Zagreb index of some complement graphs, *Adv. Math. Sci. J.* **9** (2020) 3631-3642.

[40] M. Alsharafi, A. Alameri, Y. Zeren, M. Shubatah, A. Alwardi, The Y-Index of some complement graph structures and their applications to nanotubes and nanotorus, *J. Math.* **2024** (2024) 4269325.

[41] A. Alameri, M. Alsharafi, E. Ali, M. S. Gumaan, A note on topological indices and coindices of disjunction and symmetric difference of graphs, *Discrete Appl. Math.* **304** (2021) 230–235. https://doi.org/10.1016/j.dam.2021.07.037

[42] S. A. U. H. B. Adnan, G. Abbas, T. Iqbal, Degree-based topological indices and QSPR analysis of antituberculosis drugs, *J. Chem.* **2022** (2022) 9570305.

[43] D. Balasubramaniyan, N. Chidambaram, On some neighbourhood degree-based topological indices with QSPR analysis of asthma drugs, *Eur. Phys. J. Plus* **138** (2023) 823.

[44] L. R. M. Gnanaraj, D. Ganesan, M. K. Siddiqui, Topological indices and QSPR analysis of NSAID drugs, *Polycyclic Aromat. Compd.* (2023) 1-17.

[45] M. Kalaimathi, B. J. Balamurugan, Topological indices of molecular graphs of monkeypox drugs for QSPR analysis to predict physicochemical and ADMET properties, *Int. J. Quantum Chem.* **123** (2023) e27210.

[46] N. Kansal, P. Garg, O. Singh, Temperature-based topological indices and QSPR analysis of COVID-19 drugs, *Polycyclic Aromat. Compd.* **43** (2023) 4148-4169.

[47] S. Hayat, S. J. F. Alenazi, J.-B. Liu, Two novel temperature-based topological indices with strong potential to predict physicochemical properties of polycyclic aromatic hydrocarbons with applications to silicon carbide nanotubes, *Physica Scripta* **99** (2024) 055027.

[48] S. Hayat, J.-B. Liu, Comparative analysis of temperature-based graphical indices for correlating the total π-electron energy of benzenoid hydrocarbons, *International Journal of Modern Physics B*, **39**(7) (2025) 2550047.

[49] R. Manonmani, P. Selvarani, Multilinear regression analysis of antiviral medications with topological indices, *J. Mol. Graph. Model.* **140** (2025) 109104.

[50] S. A. U. H. Bokhary, Adnan, M. K. Siddiqui, M. Cancan, On topological indices and QSPR analysis of drugs used for the treatment of breast cancer, *Polycyclic Aromat. Compd.* **42** (2021) 6233-6253.

[51] S. Hayat, M. Adil Khan, A. Khan, H. Jamil, M. Yasir Hayat Malik, Extremal hyper-Zagreb index of trees of given segments with applications to regression modeling in QSPR studies, *Alexandria Engineering Journal*, **80** (2023) 259–268.

[52] S. Hayat, M. Arshad, A. Khan, Graphs with given connectivity and their minimum Sombor index having applications to QSPR studies of monocarboxylic acids, *Heliyon*, **10** (2024) e23392.

ARTICLE IN PRESS

ARTICLE IN PRESS

[53] M. Arockiaraj, T. Rajendran, A. Jeya Shalini, K. Balasubramanian, and coauthors, Graph theoretical characterization and QSPR modeling of aluminum and magnesium hydroxide networks through integrated face degree topological indices, *Eur. Phys. J. E* **49** (2026) 14. https://doi.org/10.1140/epje/s10189-025-00554-8

[54] Y. Kara, Y. Sağlam Özkan, A. B. Bektaş, M. Arockiaraj, and coauthors, Applications of Sombor topological indices and entropy measures for QSPR modeling of anticancer drugs: A Python-based methodology, *Sci. Rep.* **16** (2026) 3005. https://doi.org/10.1038/s41598-025-32906-x

[55] G. E. Henry, S. Chandrasekaran, C. Hariharan, M. Arockiaraj, A. V. Swaminathan, Combined QSAR/QSPR, cluster modelling and docking studies of 1,3,4-oxadiazoles as anticancer drugs, *Chem. Pap.* (2026). https://doi.org/10.1007/s11696-026-04689-6

[56] M. Arockiaraj, S. Prabhu, M. Arulperumjothi, S. R. J. Kavitha, K. Balasubramanian, Topological characterization of hexagonal and rectangular tessellations of kekulenes as traps for toxic heavy metal ions, *Theor. Chem. Acc.* **140** (2021) 43. https://doi.org/10.1007/s00214-021-02733-0

[57] A. Yang, S. Sun, L. Qi, Z. Y. Kong, J. Sunarso, W. Shen, Development of an interpretable QSPR model to predict the octanol-water partition coefficient based on three artificial intelligence algorithms, *Green Chemical Engineering* **6** (2025) 193–199. https://doi.org/10.1016/j.gce.2024.07.003

[58] H. Qin, A. F. Hashem, M. F. Hanif, O. A. Fiidow, Graph theoretic and machine learning approaches in molecular property prediction of bladder cancer therapeutics, *Scientific Reports* **15** (2025) 28025. https://doi.org/10.1038/s41598-025-14175-w

[59] A. Rawat et al., Development of machine learning models to identify potentially active compounds against tuberculosis, *Scientific Reports* **15** (2025). https://doi.org/10.1038/s41598-025-04668-z

[60] J. Shi, C. Wang, L. Ni, W. Zhao, X. Yuan, Machine learning-driven QSAR modeling for predicting short-term exposure limits of hydrocarbons and their derivatives, *Processes* **13** (2025) 4025. https://doi.org/10.3390/pr13124025

[61] D. B. West, *Introduction to Graph Theory*, 2nd ed., Prentice Hall, Upper Saddle River, 2001.

[62] J. A. Bondy, U. S. R. Murty, *Graph Theory*, Springer, New York, 2008.

[63] H. Abdo, D. Dimitrov, T. Réti, D. Stevanović, Estimating the spectral radius of a graph by the second Zagreb index, *MATCH Commun. Math. Comput. Chem.* **72** (2014) 741-751.

[64] A. Golbraikh, A. Tropsha, Predictive QSAR modeling based on diversity sampling of experimental datasets for the training and test set selection, *J. Comput.-Aided Mol. Des.* **16** (2002) 357-369. https://doi.org/10.1023/A:1020869118689

[65] N. Chirico, P. Gramatica, Real external predictivity of QSAR models. Part 2. New simple metrics for the validation and comparison of predictive QSAR models, *J. Chem. Inf. Model.* **52** (2012) 2044-2058. https://doi.org/10.1021/ci300084j

[66] A. Tropsha, Best Practices for QSAR Model Development, Validation, and Exploitation, *Mol. Inf.* **29** (2010) 476-488. https://doi.org/10.1002/minf.201000061

[67] V. Kumar, S. Das, On structure sensitivity and chemical applicability of some novel degree-based topological indices, *MATCH Commun. Math. Comput. Chem.* **92** (2024) 165–203. https://doi.org/10.46793/match.92-1.165K

[68] S. Zaman, H. S. A. Yaqoob, A. Ullah, M. Sheikh, QSPR analysis of some novel drugs used in blood cancer treatment via degree based topological indices and regression models, *Polycyclic Aromat. Compd.* **44** (2024) 2458–2474. https://doi.org/10.1080/10406638.2023.2217990

ARTICLE IN PRESS