Scientific Reports

https://doi.org/10.1038/s41598-026-40969-7

Article in Press

From graph theory to chemoinformatics: modified bond-based indices and a hypothesis-driven multi-task QSAR/QSPR benchmark

Received: 26 November 2025

Accepted: 17 February 2026

Published online: 21 February 2026

Cite this article as: Altairi A., Alhaj Z., Alsharafi M. *et al.* From graph theory to chemoinformatics: modified bond-based indices and a hypothesis-driven multi-task QSAR/QSPR benchmark. *Sci Rep* (2026). https://doi.org/10.1038/s41598-026-40969-7

Azzam Altairi, Zaied Alhaj, 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

From Graph Theory to Chemoinformatics: Modified Bond-Based Indices and a Hypothesis-Driven Multi-Task QSAR/QSPR Benchmark

Azzam Altairi[1], Zaied Alhaj[1,2,⋆], Mohammed Alsharafi[2,3,4], Yusuf Zeren[4]

[1] Department of Biomedical Engineering, Institute of Graduate Studies, Istanbul University - Cerrahpasa, Türkiye

[2]Department of Biomedical Engineering, Faculty of Engineering, University of Science and Technology, Aden, Yemen

[3]Department of Mathematics, Sana'a University, Sana'a, Yemen

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

azzo.tairi@gmail.com (A.A.), zaiedalhaj@gmail.com(Z.A.)"Corresponding Author",

alsharafi205010@gmail.com(M.A.), yzeren@yildiz.edu.tr(Y.Z.)

Abstract

Graph–theoretic degree–based descriptors play a central role in chemoinformatics and QSPR/QSAR modelling, yet most classical indices either focus purely on vertex degrees or treat bond contributions in a purely multiplicative way. In this work we introduce and systematically study a new family of *modified bond-based indices* in which each edge $uv \in E(G)$ is weighted by a local bond factor $\mathcal{T}(uv) = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2$ in the *denominator*, coupled with a vertex kernel in the numerator. This construction yields modified versions of the first and second Zagreb indices, the Forgotten and Yemen indices, several connectivity-type descriptors (product, sum, Nirmala, ABC, CAB, GA, harmonic, and misbalance prodeg), as well as Sombor- and Dharwad-type bond indices.

We first present a unified edge–partition representation for any symmetric kernel, expressing each modified index as a finite sum over degree classes $E_{(a,b)}(G)$. This framework allows us to derive closed-form expressions for all sixteen modified bond-based indices on a broad collection of benchmark families: paths P_{n}, cycles C_{n}, complete graphs K_{n}, complete bipartite graphs $K_{m,\mathrm{n}}$, stars S_{n}, friendship graphs F_{n}, wheels W_{n}, book graphs B_{n}, Dutch windmill graphs $D_{\mathrm{n}}^{(m)}$, and hypercubes Q_d. The resulting tables reveal clear asymptotic growth patterns and highlight which structures are extremal for the modified descriptors.

Moreover, we obtain sharp degree–extreme bounds for a representative subset of the indices in terms of the order n, size m, and the minimum and maximum degrees δ and Δ, with equality characterizing regular graphs. The proposed modified bond-based indices thus provide a flexible and analytically tractable family of descriptors that couple vertex and bond information in a novel way, and are well suited as structured features for modern chemoinformatics and graph-based machine-learning models on molecular graphs. Finally, to demonstrate predictive utility in a hypothesis-driven setting, we further benchmark these ^{m}BI descriptors within a large multi-task QSAR/QSPR pipeline on 3,219 ChEMBL antibacterial molecules across ten continuous properties using a heterogeneous model zoo under three descriptor scenarios, where the combined descriptors scenario achieves the best overall generalisation (Macro Test $R^2 = 0.861$; Global zRMSE = 0.373), improving upon the Physicochemical descriptors scenario (Macro Test $R^2 = 0.852$; Global zRMSE = 0.385).

Keywords: Modified bond-based indices, molecular graph theory, physicochemical properties, QSPR, QSAR, antibacterial, drug discovery

1 Introduction and Literature Review

Chemical graph theory models a molecule as a graph whose vertices represent atoms and whose edges represent chemical bonds, with the goal of extracting graph invariants that encode chemically meaningful structural information [1, 2]. Early and still widely used distance–based descriptors include the Wiener index introduced in [3] and the Hosoya index introduced in [4]. In parallel,

ARTICLE IN PRESS

degree–based descriptors were developed as predictors of physicochemical behaviour: the Zagreb indices emerged from the molecular-orbital viewpoint in [5] and were later surveyed in [6, 7], with further systematic coverage in [8]. The Randić connectivity index was proposed in [9] and its subsequent development and applications are discussed in [10, 11]. Comprehensive treatments of molecular descriptors and their use within QSPR/QSAR pipelines can be found in [12, 13], while a recent applied study using line-graph descriptors is reported in [14]. Classical QSPR examples include modelling boiling points of haloalkanes in [15], the PI descriptor proposed in [16], and the atom–bond connectivity (ABC) index for enthalpies of formation introduced in [17].

A very large family of topological indices has been proposed, with degree–based descriptors playing a prominent role. Structural and extremal properties of Zagreb-type indices were developed in [18, 19], while more recent sharp bounds and Nordhaus–Gaddum type results are given in [20, 21]. For Randić-type descriptors, edited and book-length resources include [22] and [23]. Additional degree-based indices include the geometric–arithmetic (GA) index introduced in [24] and further studied in [25, 26], as well as the harmonic index introduced in [27] with representative bounds in [28] and a broader survey in [29]. The ABC index was further analysed from an energetic perspective in [30] and investigated for trees in [31]; a related connectivity construction was proposed in [32], and a review of extremal results and bounds for ABC-type quantities is provided in [33]. General sum-connectivity indices were studied in [34], while extremal problems for broad classes of degree-based indices were treated in [35] and a geometric viewpoint was developed in [36]; applications of generalized Zagreb indices are discussed in [37].

Indices emphasising squared degrees form another influential line of work. The forgotten (or F-) index was introduced in [38] and subsequent bounds were established in [39, 40]. Inequalities for general Zagreb-type quantities appear in [41], and extremal trees with respect to the F-index were investigated in [42]. Variants and computations for special families include thorny graphs in [43] and nanostructure-related settings in [44], while drug-structure applications are reported in [45].

To increase sensitivity to local structural changes beyond simple degrees, neighbourhood degree–based and eccentric neighbourhood constructions have also been introduced. General neighbourhood degree-based indices were proposed in [46], and neighbourhood Zagreb indices for product graphs were studied in [47]. Eccentric neighbourhood Zagreb indices and their computational significance were analysed in [48], and chemically motivated neighbourhood inverse-sum indeg indices were investigated in [49]. Related structural constraints for neighbourhood degree lists (in diameter-2 graphs) are studied in [50], and an example of QSPR analysis for asthma drugs using neighbourhood degree-type indices is presented in [51].

Recently, further descriptors have attracted attention due to both extremal theory and regression performance. Extremal hyper-Zagreb results with QSPR-oriented regression modelling are reported in [52]. Bounds and extremal behaviour for broader families of indices are discussed in [53, 54], and irregularity-type developments are treated in [55, 56]. The Sombor index has motivated additional structural investigations, including connectivity-constrained minimization with QSPR applications in [57]. Temperature-based indices with QSPR potential are proposed in [58], and a comparative study of temperature-based graphical indices for benzenoid hydrocarbons is given in [59].

Of particular relevance to the present manuscript are edge-sum descriptors that incorporate endpoint degrees and explicitly treat bonds as primary carriers of information. The symmetric division degree index was introduced in [60] and assessed as a molecular descriptor in [61], while degree-extreme bounds were derived in [62, 63]. Inverse variants and related results were developed in [64, 65], and a general reduced second Zagreb-type quantity was studied in [66]; additional

ARTICLE IN PRESS

results for symmetric division degree constructions appear in [67]. Bond-additive modelling was formalized through Adriatic indices in [68] and further developed for rodeg-type quantities in [69]. A GA/AM-ratio edge index was introduced in [70], and more recent bond-incident degree results for chemical trees with fixed leaves are given in [71]. Atom–bond sum-connectivity indices were introduced in [72] and extended to unicyclic graphs with applications in [73].

Beyond degree-based constructions, spectral and distance-based descriptors enrich the landscape. The Estrada index was introduced for characterizing molecular structure in [74], studied further in [75], and analysed on alkanes with extreme behaviour in [76]. Computations of multiple topological indices on nanostructure frameworks are presented in [77], and the Padmakar–Ivan index for multi-walled nanotubes/nanotori is treated in [78]. Complement-graph perspectives and additional descriptors for nanostructures are discussed in [79, 80], and related coindex/topological considerations are illustrated in [81].

From an applied QSAR/QSPR perspective, degree-based and related indices have repeatedly produced competitive predictive models for pharmacologically relevant compounds. Antituberculosis and NSAID studies are reported in [82, 83], while modelling of breast-cancer and antiviral datasets appears in [84, 85]. Applications to emerging therapeutic targets include COVID-19-related modelling in [86] and monkeypox drug graphs in [87].

ARTICLE IN PRESS

Gap and contribution. The literature above provides a rich collection of vertex-centric and bond-centric descriptors, but comparatively fewer works study a *unified mechanism* that systematically modulates a classical symmetric kernel in endpoint degrees by a local bond-normalization factor. This motivates the present paper: we develop a coherent framework of modified bond-based (edge-degree–modulated) indices that (i) preserves the interpretability of standard degree-based kernels, (ii) incorporates explicit bond-degree normalization to enhance local sensitivity, (iii) supports rigorous graph-theoretic analysis (including degree-extreme bounds on relevant families), and (iv) enables validation via a multi-task QSAR/QSPR study on a large antibacterial dataset.

These observations suggest that new families of structurally informed indices—such as the modified bond-based descriptors introduced here—are natural candidates for integration into modern chemoinformatics and QSPR/QSAR workflows.

Building on this motivation, we move beyond purely analytical development and subject the proposed modified bond-based indices (^{m}BI) to a hypothesis-driven predictive evaluation. Specifically, we test whether ^{m}BI descriptors contribute complementary information relative to standard physicochemical descriptors in realistic QSAR/QSPR settings, rather than acting as redundant reparameterizations. To this end, we construct a multi-task regression evaluation on a curated ChEMBL antibacterial dataset with ten continuous endpoints and compare three descriptor scenarios: ^{m}BI alone, physicochemical descriptors alone, and their concatenation. We consider a diverse model suite spanning linear, kernel-based, robust, and ensemble/boosting regressors. Using a fixed molecule-level train–test split, leakage control when targets overlap with predictors, and consistent cross-validation and test evaluation across all endpoints, the resulting performance patterns provide a direct empirical test of the central claim: whether incorporating ^{m}BI indices yields measurable gains in generalization at both per-endpoint and aggregate multi-target levels.

ARTICLE IN PRESS

Chemoinformatics perspective on modified bond-based indices

Classical degree-based molecular descriptors encode structural information at the level of *vertices* (atoms). Typical examples include the Zagreb, forgotten and Sombor indices, which aggregate functions of the vertex degrees over the whole graph. By contrast, edge-centric reformulations emphasise *bonds* and treat edges as the primary carriers of information. The modified bond-based indices introduced in this work are designed to merge these two viewpoints within a single, coherent framework.

More precisely, each edge $uv \in E(G)$ is assigned a weight of the form

$$\frac{\Psi\big(\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\big)}{\mathcal{T}(uv)}, \qquad \mathcal{T}(uv) = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2,$$

where the *bond factor* $\mathcal{T}(uv)$ quantifies the local bond environment of uv, and the *vertex kernel* $\Psi(\phi_G(\mathrm{u}), \phi_G(\mathrm{v}))$ encodes the valence pattern of the incident atoms (see Figure 1). This coupled edge–vertex encoding is sensitive to subtle local rearrangements, such as isomeric substitutions that modify edge neighbourhoods while leaving the global degree sequence unchanged or nearly unchanged. Moreover, by choosing Ψ appropriately, one can recover modified analogues of many well-known degree-based indices, which facilitates direct comparison with the existing chemoinformatics literature.

ARTICLE IN PRESS

Pendant edges and chemical interpretation of the bond factor. In molecular graphs, pendant (terminal) atoms are common (e.g., terminal substituents or terminal heavy atoms in acyclic fragments) and correspond to vertices of degree 1. Such pendant edges are *not excluded* by our framework. Indeed, if uv is an edge with $\phi_G(u) = 1$ and $\phi_G(v) \geq 2$, then

$$\mathcal{T}(uv) = \phi_G(u) + \phi_G(v) - 2 = \phi_G(v) - 1 \geq 1,$$

so the edge contribution $\Psi(\phi_G(u), \phi_G(v))/\mathcal{T}(uv)$ remains finite and well-defined. Chemically, the normalization by $\mathcal{T}(uv)$ down-weights bonds incident to highly connected atoms (locally crowded bonding environments) and gives relatively higher weight to bonds in less crowded local environments, while still allowing terminal bonds to contribute in a controlled, interpretable way. The only case in which $\mathcal{T}(uv) = 0$ occurs is when *both* endpoints have degree 1, i.e., the connected graph $G \cong K_2$, which does not arise in standard molecular datasets used for QSAR/QSPR (and, in particular, does not occur in our benchmark). Therefore, no special preprocessing or exclusion of pendant edges is required for practical chemoinformatics use of the proposed indices.

From a chemoinformatics and QSPR/QSAR perspective, the modified bond-based indices therefore provide a flexible family of local, interpretable descriptors that naturally respect the atom–bond structure of molecular graphs. Their strictly local definition makes them suitable both as stand-alone predictors and as engineered features within graph-based machine-learning models, where they can complement modern message-passing schemes by injecting chemically meaningful, degree-aware information at the bond level.

ARTICLE IN PRESS

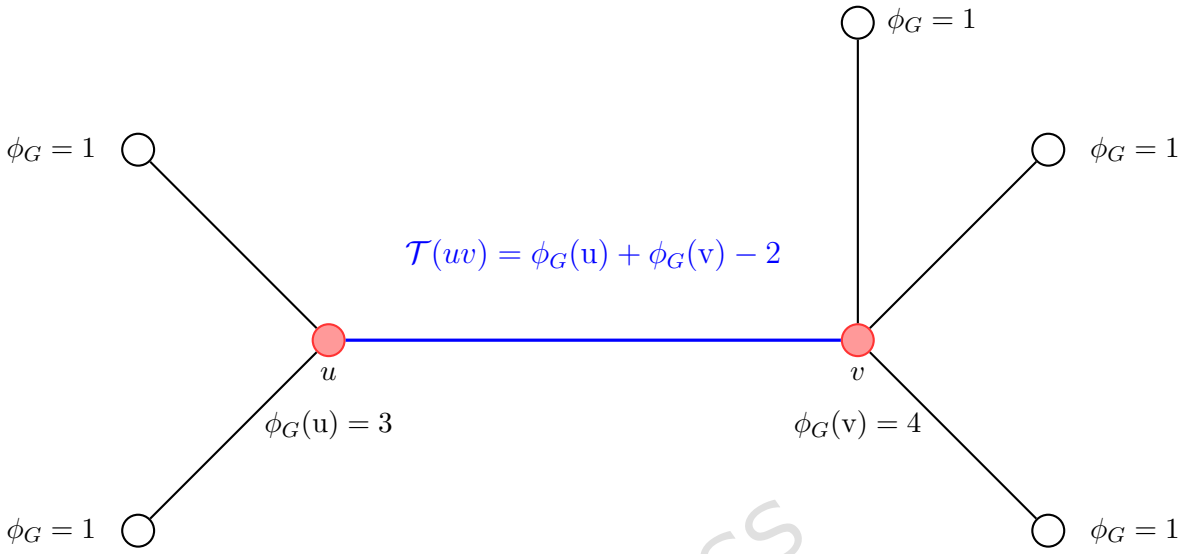

Figure 1: Schematic atom–bond representation of an edge uv in a molecular graph G. The vertex degrees $\phi_G(\mathrm{u})$ and $\phi_G(\mathrm{v})$ and the associated bond factor $\mathcal{T}(uv) = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2$ are indicated. The vertices u and v are shown in red and the bond uv in blue.

2 Properties of Modified Bond-Based Degree Indices

Throughout this section, let $G = (V, E)$ be a simple, finite, connected, undirected graph. For $\mathrm{u} \in V$, write $\phi_G(\mathrm{u})$ for its degree. For an edge $e = uv \in E$, the (edge/bond) degree is defined by

$$\mathcal{T}(e) \;=\; \mathcal{T}(uv) \;:=\; \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2.$$

This counts the number of edges incident to u or v other than uv itself; in particular, $\mathcal{T}(uv) = 0$ if and only if $\phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 1$ (i.e., $G \cong K_2$ in the connected case). In particular, $\mathcal{T}(uv) > 0$ for every edge whenever G is connected and $|V(G)| \geq 3$.

In what follows we assume $\mathcal{T}(uv) > 0$ for all edges $uv \in E(G)$ (equivalently, $G \ncong K_2$ in the connected case). For the degree–extreme bounds we additionally assume $\delta(G) \geq 2$. (In particular, pendant edges with a degree-1 endpoint are permitted and satisfy $\mathcal{T}(uv) \geq 1$ whenever $|V(G)| \geq 3$.)

We also introduce the abbreviations

$$\mathcal{S}_{uv} = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}), \qquad \mathcal{P}_{uv} = \phi_G(\mathrm{u})\,\phi_G(\mathrm{v}), \qquad \mathcal{T}(uv) = \mathcal{S}_{uv} - 2.$$

The quantity $\mathcal{T}(uv)$ counts the number of edges incident with u or v other than uv itself; note that for connected graphs with $|V(G)| \geq 3$ one has $\mathcal{T}(uv) \geq 1$ for all $uv \in E(G)$.

Definition 1. *Let $\Psi : \mathbb{N} \times \mathbb{N} \to \mathbb{R}^+$ be a symmetric kernel in the endpoint degrees. The* modified bond-based *(edge-degree–modulated) analogue induced by Ψ is*

$${}^{m}\mathrm{BI}_{\Psi}(G) \;:=\; \sum_{uv \in E(G)} \frac{\Psi\big(\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\big)}{\mathcal{T}(uv)} \;=\; \mathcal{B}_{\Psi}(G).$$

(We assume $\Psi(\phi_G(u), \phi_G(v)) > 0$ for all $uv \in E(G)$. For connected graphs with $|V(G)| \geq 3$, one has $\mathcal{T}(uv) \geq 1$ for every edge, hence ${}^{m}\mathrm{BI}_{\Psi}(G)$ is well-defined.)

ARTICLE IN PRESS

For connected graphs with $|V(G)| \geq 3$, we have $\mathcal{T}(uv) \geq 1$ for all edges $uv \in E(G)$, hence the indices are well-defined.

Choosing Ψ to be the kernels of well-known degree-based indices yields the following modified bond-based instances.

Modified bond Zagreb indices.

$$
\begin{aligned}
{}^{m}\mathrm{BM}_1(G) &= \sum_{uv\in E(G)} \frac{\phi_G(\mathrm{u}) + \phi_G(\mathrm{v})}{\mathcal{T}(uv)} = \sum_{uv\in E(G)} \frac{\mathcal{S}_{uv}}{\mathcal{T}(uv)}, && \big(\Psi(x,y) = x + y\big),\\
{}^{m}\mathrm{BM}_2(G) &= \sum_{uv\in E(G)} \frac{\phi_G(\mathrm{u})\,\phi_G(\mathrm{v})}{\mathcal{T}(uv)} = \sum_{uv\in E(G)} \frac{\mathcal{P}_{uv}}{\mathcal{T}(uv)}, && \big(\Psi(x,y) = xy\big).
\end{aligned}
$$

Modified bond Forgotten and Yemen indices.

$$
\begin{aligned}
{}^{m}\mathrm{BF}(G) &= \sum_{uv\in E(G)} \frac{\phi_G(\mathrm{u})^2 + \phi_G(\mathrm{v})^2}{\mathcal{T}(uv)} && \big(\Psi(x,y) = x^2 + y^2\big),\\
{}^{m}\mathrm{BY}(G) &= \sum_{uv\in E(G)} \frac{\phi_G(\mathrm{u})^3 + \phi_G(\mathrm{v})^3}{\mathcal{T}(uv)} && \big(\Psi(x,y) = x^3 + y^3\big).
\end{aligned}
$$

Modified bond product–connectivity (Randić-type).

$$
{}^{m}\mathrm{BPC}(G) = \sum_{uv\in E(G)} \frac{1}{\mathcal{T}(uv)\sqrt{\phi_G(\mathrm{u})\,\phi_G(\mathrm{v})}} = \sum_{uv\in E(G)} \frac{1}{\mathcal{T}(uv)\sqrt{\mathcal{P}_{uv}}}, \quad \big(\Psi(x,y) = (xy)^{-1/2}\big).
$$

Modified bond sum–connectivity family.

$$
\begin{aligned}
{}^{m}\mathrm{BSC}(G) &= \sum_{uv\in E(G)} \frac{1}{\mathcal{T}(uv)\sqrt{\phi_G(\mathrm{u}) + \phi_G(\mathrm{v})}} = \sum_{uv\in E(G)} \frac{1}{\mathcal{T}(uv)\sqrt{\mathcal{S}_{uv}}}, && \big(\Psi(x,y) = (x+y)^{-1/2}\big),\\
{}^{m}\mathrm{BN}(G) &= \sum_{uv\in E(G)} \frac{\sqrt{\phi_G(\mathrm{u}) + \phi_G(\mathrm{v})}}{\mathcal{T}(uv)} = \sum_{uv\in E(G)} \frac{\sqrt{\mathcal{S}_{uv}}}{\mathcal{T}(uv)} && \big(\Psi(x,y) = \sqrt{x+y}\big).
\end{aligned}
$$

Modified bond atom–bond connectivity (ABC) and its reciprocal form.

$$
\begin{aligned}
{}^{m}\mathrm{BABC}(G) &= \sum_{uv\in E(G)} \frac{1}{\mathcal{T}(uv)}\sqrt{\frac{\phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2}{\phi_G(\mathrm{u})\,\phi_G(\mathrm{v})}} = \sum_{uv\in E(G)} \frac{1}{\sqrt{\mathcal{P}_{uv}\,\mathcal{T}(uv)}}, && \big(\Psi(x,y) = \sqrt{\tfrac{x+y-2}{xy}}\big),\\
{}^{m}\mathrm{BCAB}(G) &= \sum_{uv\in E(G)} \frac{1}{\mathcal{T}(uv)}\sqrt{\frac{\phi_G(\mathrm{u})\,\phi_G(\mathrm{v})}{\phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2}} = \sum_{uv\in E(G)} \frac{\sqrt{\mathcal{P}_{uv}}}{\mathcal{T}(uv)^{3/2}}, && \big(\Psi(x,y) = \sqrt{\tfrac{xy}{x+y-2}}\big).
\end{aligned}
$$

Modified bond Gourava indices.

$$
\begin{aligned}
{}^{m}\mathrm{BGO}_1(G) &= \sum_{uv\in E(G)} \frac{\phi_G(\mathrm{u})\,\phi_G(\mathrm{v}) + \phi_G(\mathrm{v})\,\phi_G(\mathrm{u})}{\mathcal{T}(uv)} = \sum_{uv\in E(G)} \frac{2\mathcal{P}_{uv}}{\mathcal{T}(uv)}, && \big(\Psi(x,y) = 2xy\big),\\
{}^{m}\mathrm{BGO}_2(G) &= \sum_{uv\in E(G)} \frac{\phi_G(\mathrm{u})^2\,\phi_G(\mathrm{v}) + \phi_G(\mathrm{v})^2\,\phi_G(\mathrm{u})}{\mathcal{T}(uv)} = \sum_{uv\in E(G)} \frac{\mathcal{P}_{uv}\,\mathcal{S}_{uv}}{\mathcal{T}(uv)}, && \big(\Psi(x,y) = x^2y + y^2x\big).
\end{aligned}
$$

ARTICLE IN PRESS

Modified bond geometric–arithmetic (GA).

$$ {}^{m}\mathrm{BGA}(G) \;=\; \sum_{uv\in E(G)} \frac{2\sqrt{\phi_G(\mathrm{u})\,\phi_G(\mathrm{v})}}{\phi_G(\mathrm{u})+\phi_G(\mathrm{v})}\,\frac{}{\mathcal{T}(uv)} \;=\; \sum_{uv\in E(G)} \frac{2\sqrt{\mathcal{P}_{uv}}}{\mathcal{T}(uv)\,\mathcal{S}_{uv}}, \quad \big(\Psi(x,y)=\tfrac{2\sqrt{xy}}{x+y}\big). $$

Modified bond harmonic (H).

$$ {}^{m}\mathrm{BH}(G) \;=\; \sum_{uv\in E(G)} \frac{2}{\phi_G(\mathrm{u})+\phi_G(\mathrm{v})}\,\frac{}{\mathcal{T}(uv)} \;=\; \sum_{uv\in E(G)} \frac{2}{\mathcal{T}(uv)\,\mathcal{S}_{uv}}, \quad \big(\Psi(x,y)=\tfrac{2}{x+y}\big). $$

Modified bond misbalance prodeg (MPI-type).

$$ {}^{m}\mathrm{BMP}(G) \;=\; \sum_{uv\in E(G)} \frac{\sqrt{\phi_G(\mathrm{u})}+\sqrt{\phi_G(\mathrm{v})}}{\mathcal{T}(uv)} \big(\Psi(x,y)=\sqrt{x}+\sqrt{y}\big). $$

Modified bond Sombor (SO) and Dharwad-type.

$$ {}^{m}\mathrm{BS}(G) \;=\; \sum_{uv\in E(G)} \frac{\sqrt{\phi_G(\mathrm{u})^2+\phi_G(\mathrm{v})^2}}{\mathcal{T}(uv)} \qquad \big(\Psi(x,y)=\sqrt{x^2+y^2}\big), $$

$$ {}^{m}\mathrm{BD}(G) \;=\; \sum_{uv\in E(G)} \frac{\sqrt{\phi_G(\mathrm{u})^3+\phi_G(\mathrm{v})^3}}{\mathcal{T}(uv)} \qquad \big(\Psi(x,y)=\sqrt{x^3+y^3}\big). $$

ARTICLE IN PRESS

2.1 Relations to Classical Degree-Based Indices

For a fixed symmetric kernel $\Psi : \mathbb{N}\times\mathbb{N}\to\mathbb{R}^{+}$, it is convenient to introduce the corresponding *classical edge-degree index*

$$ I_{\Psi}(G) \;:=\; \sum_{uv\in E(G)} \Psi\big(\phi_G(\mathrm{u}),\phi_G(\mathrm{v})\big), \tag{1} $$

so that the modified bond-based analogue is

$$ {}^{m}\mathrm{BI}_{\Psi}(G) \;:=\; \sum_{uv\in E(G)} \frac{\Psi\big(\phi_G(\mathrm{u}),\phi_G(\mathrm{v})\big)}{\mathcal{T}(uv)}, \qquad \mathcal{T}(uv)=\phi_G(\mathrm{u})+\phi_G(\mathrm{v})-2. \tag{2} $$

Thus each modified bond-based index associated with a kernel Ψ is obtained from its classical counterpart by reweighting each edge term by the inverse bond-degree factor $1/\mathcal{T}(uv)$.

Remark 2.1. *Let*

$$ \rho_{\min}(G) := \min_{uv\in E(G)} \mathcal{T}(uv), \qquad \rho_{\max}(G) := \max_{uv\in E(G)} \mathcal{T}(uv) $$

be the extremal bond degrees of G. Then $\rho_{\min}(G)\ge 0$, and $\rho_{\min}(G)>0$ whenever G has no pendant edges. When G is k-regular, one has $\mathcal{T}(uv)=2k-2$ for all $uv\in E(G)$.

ARTICLE IN PRESS

Proposition 2.2. *Let G be a connected graph without pendant edges (so that $\mathcal{T}(uv) \geq 1$ for all $uv \in E(G)$), and let Ψ be a symmetric kernel as above. Then*

$$\frac{1}{\rho_{\max}(G)}\, I_\Psi(G) \;\leq\; {}^m\mathrm{BI}_\Psi(G) \;\leq\; \frac{1}{\rho_{\min}(G)}\, I_\Psi(G).$$

Proof. For each edge $uv \in E(G)$,

$$\frac{1}{\rho_{\max}(G)} \;\leq\; \frac{1}{\mathcal{T}(uv)} \;\leq\; \frac{1}{\rho_{\min}(G)}.$$

Multiplying by $\Psi(\phi_G(\mathrm{u}), \phi_G(\mathrm{v})) > 0$ and summing over all edges gives the claimed inequality. □

Corollary 2.3. *If G is k-regular, then $\mathcal{T}(uv) = 2k - 2$ for all edges $uv \in E(G)$ and*

$${}^m\mathrm{BI}_\Psi(G) \;=\; \frac{1}{2k-2}\, I_\Psi(G).$$

In particular, for any k-regular graph G the modified bond-based indices are simply rescaled versions of their classical edge-degree analogues.

Corollary 2.4. *Let Ψ be a symmetric kernel.*

- *For the cycle C_n ($\mathrm{n} \geq 3$), $G = C_\mathrm{n}$ is 2-regular, so*

$${}^m\mathrm{BI}_\Psi(C_\mathrm{n}) \;=\; \frac{1}{2}\, I_\Psi(C_\mathrm{n}) \;=\; \frac{\mathrm{n}}{2}\, \Psi(2,2).$$

- *For the complete graph K_n ($\mathrm{n} \geq 3$), $G = K_\mathrm{n}$ is $(\mathrm{n}-1)$-regular with $|E(K_\mathrm{n})| = \frac{\mathrm{n}(\mathrm{n}-1)}{2}$. Moreover, for every edge $uv \in E(K_n)$ we have*

$$\mathcal{T}(uv) = \phi_{K_n}(u) + \phi_{K_n}(v) - 2 = 2(\mathrm{n}-1) - 2 = 2\mathrm{n} - 4 = 2(\mathrm{n}-2),$$

  so the regular normalization factor equals $\frac{1}{2\mathrm{n}-4}$. Hence

$$\begin{aligned} {}^m\mathrm{BI}_\Psi(K_\mathrm{n}) \;&=\; \frac{1}{2\mathrm{n}-4}\, I_\Psi(K_\mathrm{n}) \;=\; \frac{1}{2\mathrm{n}-4} \cdot \frac{\mathrm{n}(\mathrm{n}-1)}{2}\, \Psi(\mathrm{n}-1, \mathrm{n}-1) \\ &=\; \frac{\mathrm{n}(\mathrm{n}-1)}{4(\mathrm{n}-2)}\, \Psi(\mathrm{n}-1, \mathrm{n}-1). \end{aligned}$$

- *For the d-dimensional cube Q_d ($d \geq 1$), $G = Q_d$ is d-regular with 2^d vertices and $|E(Q_d)| = d\, 2^{d-1}$, hence*

$${}^m\mathrm{BI}_\Psi(Q_d) \;=\; \frac{1}{2d-2}\, I_\Psi(Q_d) \;=\; \frac{d\, 2^{d-1}}{2d-2}\, \Psi(d,d) \;=\; \frac{d\, 2^{d-1}}{2(d-1)}\, \Psi(d,d).$$

Remark 2.5. *Corollary 2.4 shows that on regular benchmark families (cycles, complete graphs, hypercubes) the passage from classical degree-based indices to their modified bond-based versions amounts only to a simple normalisation factor depending on the common degree. In non-regular families, the edge-partition viewpoint below becomes more informative.*

ARTICLE IN PRESS

ARTICLE IN PRESS

2.2 Edge-partition representation and standard graph families

For many structured graphs it is natural to partition the edge set according to the degrees of their endpoints.

For fixed integers $a, b \geq 1$ define

$$E_{(a,b)}(G) \;:=\; \big\{\, uv \in E(G) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{a, b\} \,\big\},$$

and let $|E_{(a,b)}(G)|$ denote the cardinality of this subset.

Proposition 2.6. *Let G be a graph such that $\mathcal{T}(uv) > 0$ for every edge $uv \in E(G)$ (for instance, any graph without edges joining two pendant vertices). Let $\Psi : \mathbb{N} \times \mathbb{N} \to \mathbb{R}^+$ be a symmetric kernel. Then*

$$^{m}\mathrm{BI}_{\Psi}(G) \;=\; \sum_{(a,b)} \frac{\Psi(a,b)}{a+b-2}\, |E_{(a,b)}(G)|,$$

where the sum ranges over all unordered degree pairs (a, b) that occur among the edges of G.

Proof. For each edge $uv \in E(G)$ with $\{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{a, b\}$ we have

$$\mathcal{T}(uv) = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2 = a + b - 2,$$

and the corresponding contribution

$$\frac{\Psi(\phi_G(\mathrm{u}), \phi_G(\mathrm{v}))}{\mathcal{T}(uv)} = \frac{\Psi(a,b)}{a+b-2}$$

is constant on $E_{(a,b)}(G)$. Summing over $E_{(a,b)}(G)$ and then over all degree pairs (a, b) gives the claimed expression. □

We now record the relevant degree partitions for the benchmark families considered in this paper. Explicit closed forms for the modified indices $^{m}\mathrm{BM}_1, \ldots, {}^{m}\mathrm{BD}$ follow by direct substitution into Proposition 2.6 and are collected in the subsequent subsection.

Path graph P_{n}

For the path P_{n} ($\mathrm{n} \geq 3$), the degrees are

$$\phi_G(\mathrm{v}) = \begin{cases} 1, & v \in \{v_1, v_n\}, \\ 2, & \text{otherwise}, \end{cases}$$

and there are two edge types:

$$|E_{(1,2)}(P_{\mathrm{n}})| = 2, \qquad |E_{(2,2)}(P_{\mathrm{n}})| = \mathrm{n} - 3,$$

with

$$\mathcal{T}(1,2) = 1, \qquad \mathcal{T}(2,2) = 2.$$

Hence

$$^{m}\mathrm{BI}_{\Psi}(P_{\mathrm{n}}) = 2\,\Psi(1,2) + (\mathrm{n}-3)\,\frac{\Psi(2,2)}{2}.$$

ARTICLE IN PRESS

Cycle graph C_{n}

For the cycle C_{n} ($\mathrm{n} \geq 3$), all vertices have degree 2, and the edge set consists of a single degree class

$$E_{(2,2)}(C_{\mathrm{n}}) = \big\{\, uv \in E(C_{\mathrm{n}}) \;:\; \phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 2 \,\big\}, \qquad |E_{(2,2)}(C_{\mathrm{n}})| = \mathrm{n},$$

with $\mathcal{T}(2,2) = 2$. Thus

$$^{m}\mathrm{BI}_{\Psi}(C_{\mathrm{n}}) = \frac{\mathrm{n}}{2}\,\Psi(2,2),$$

in agreement with Corollary 2.4 for 2-regular graphs.

Star graph S_{n}

We adopt the convention that S_{n} is the star on $\mathrm{n} \geq 3$ vertices, i.e. $S_{\mathrm{n}} \cong K_{1,\mathrm{n}-1}$. The central vertex has degree $\mathrm{n}-1$ and the $\mathrm{n}-1$ pendant vertices have degree 1. Hence

$$E_{(1,\mathrm{n}-1)}(S_{\mathrm{n}}) = \big\{\, uv \in E(S_{\mathrm{n}}) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{1, \mathrm{n}-1\} \,\big\}, \qquad |E_{(1,\mathrm{n}-1)}(S_{\mathrm{n}})| = \mathrm{n}-1,$$

with

$$\mathcal{T}(1, \mathrm{n}-1) = 1 + (\mathrm{n}-1) - 2 = \mathrm{n}-2.$$

Therefore

$$^{m}\mathrm{BI}_{\Psi}(S_{\mathrm{n}}) = (\mathrm{n}-1)\,\frac{\Psi(1,\mathrm{n}-1)}{\mathrm{n}-2}.$$

Complete bipartite graph $K_{m,\mathrm{n}}$

For the complete bipartite graph $K_{m,\mathrm{n}}$ ($m, \mathrm{n} \geq 2$), the m vertices in one part have degree n and the n vertices in the other part have degree m. Every edge joins a vertex of degree m to a vertex of degree n:

$$E_{(m,\mathrm{n})}(K_{m,\mathrm{n}}) = \big\{\, uv \in E(K_{m,\mathrm{n}}) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{m, \mathrm{n}\} \,\big\},$$

with

$$|E_{(m,\mathrm{n})}(K_{m,\mathrm{n}})| = m\mathrm{n}, \qquad \mathcal{T}(m,\mathrm{n}) = m + \mathrm{n} - 2.$$

Thus

$$^{m}\mathrm{BI}_{\Psi}(K_{m,\mathrm{n}}) = m\mathrm{n}\,\frac{\Psi(m,\mathrm{n})}{m+\mathrm{n}-2}.$$

Friendship graph F_{n}

The friendship graph F_{n} consists of n triangles sharing a common central vertex. The central vertex has degree 2n and each of the 2n outer vertices has degree 2. The edge set decomposes as

$$E_{(2,2)}(F_{\mathrm{n}}) = \{\, uv \in E(F_{\mathrm{n}}) \;:\; \phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 2 \,\}, \qquad |E_{(2,2)}(F_{\mathrm{n}})| = \mathrm{n},$$

$$E_{(2,2\mathrm{n})}(F_{\mathrm{n}}) = \{\, uv \in E(F_{\mathrm{n}}) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{2, 2\mathrm{n}\} \,\}, \qquad |E_{(2,2\mathrm{n})}(F_{\mathrm{n}})| = 2\mathrm{n},$$

with bond degrees

$$\mathcal{T}(2,2) = 2, \qquad \mathcal{T}(2,2\mathrm{n}) = 2 + 2\mathrm{n} - 2 = 2\mathrm{n}.$$

Therefore

$$^{m}\mathrm{BI}_{\Psi}(F_{\mathrm{n}}) = \frac{\mathrm{n}}{2}\,\Psi(2,2) + \Psi(2,2\mathrm{n}).$$

ARTICLE IN PRESS

Wheel graph W_{n}. We follow the standard convention that the wheel W_{n} is obtained from the cycle C_{n} by adding a new universal vertex (the hub) adjacent to all vertices of C_{n}. Thus $|V(W_{\mathrm{n}})| = \mathrm{n}+1$, the n outer vertices have degree 3, and the hub has degree n.

The edge set $E(W_{\mathrm{n}})$ splits into

$$E_{(3,3)}(W_{\mathrm{n}}) = \{\, uv \in E(W_{\mathrm{n}}) \;:\; \phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 3 \,\}, \qquad |E_{(3,3)}(W_{\mathrm{n}})| = \mathrm{n},$$

$$E_{(3,\mathrm{n})}(W_{\mathrm{n}}) = \{\, uv \in E(W_{\mathrm{n}}) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{3, \mathrm{n}\} \,\}, \qquad |E_{(3,\mathrm{n})}(W_{\mathrm{n}})| = \mathrm{n},$$

with bond degrees

$$\mathcal{T}(3,3) = 3+3-2 = 4, \qquad \mathcal{T}(3,\mathrm{n}) = 3+\mathrm{n}-2 = \mathrm{n}+1.$$

Hence, for any symmetric kernel Ψ,

$${}^{m}\mathrm{BI}_{\Psi}(W_{\mathrm{n}}) = n\left(\frac{\Psi(3,3)}{4} + \frac{\Psi(3,\mathrm{n})}{\mathrm{n}+1}\right).$$

Book graph B_{n}. The book graph B_{n} consists of n copies of C_4 sharing a common edge. The two vertices of the common edge (the spine) each have degree $\mathrm{n}+1$, and the remaining 2n vertices each have degree 2. The edge set decomposes into

$$E_{(2,2)}(B_{\mathrm{n}}) = \{\, uv \in E(B_{\mathrm{n}}) \;:\; \phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 2 \,\}, \qquad |E_{(2,2)}(B_{\mathrm{n}})| = \mathrm{n},$$

$$E_{(2,\mathrm{n}+1)}(B_{\mathrm{n}}) = \{\, uv \in E(B_{\mathrm{n}}) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{2, \mathrm{n}+1\} \,\}, \qquad |E_{(2,\mathrm{n}+1)}(B_{\mathrm{n}})| = 2\mathrm{n},$$

$$E_{(\mathrm{n}+1,\mathrm{n}+1)}(B_{\mathrm{n}}) = \{\, uv \in E(B_{\mathrm{n}}) \;:\; \phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = \mathrm{n}+1 \,\}, \qquad |E_{(\mathrm{n}+1,\mathrm{n}+1)}(B_{\mathrm{n}})| = 1,$$

with

$$\mathcal{T}(2,2) = 2, \qquad \mathcal{T}(2,\mathrm{n}+1) = \mathrm{n}+1, \qquad \mathcal{T}(\mathrm{n}+1,\mathrm{n}+1) = 2\mathrm{n}.$$

Therefore

$${}^{m}\mathrm{BI}_{\Psi}(B_{\mathrm{n}}) = \mathrm{n}\,\frac{\Psi(2,2)}{2} + 2\mathrm{n}\,\frac{\Psi(2,\mathrm{n}+1)}{\mathrm{n}+1} + \frac{\Psi(\mathrm{n}+1,\mathrm{n}+1)}{2\mathrm{n}}.$$

Dutch windmill graph $D_{\mathrm{n}}^{(r)}$. The Dutch windmill graph $D_{\mathrm{n}}^{(r)}$ ($\mathrm{n} \geq 3$, $r \geq 2$) is obtained by identifying one vertex of r copies of C_{n}. The identified central vertex has degree $2r$, and all other vertices have degree 2. The edge set decomposes into

$$E_{(2,2)}(D_{\mathrm{n}}^{(r)}) = \{\, uv \in E(D_{\mathrm{n}}^{(r)}) \;:\; \phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 2 \,\}, \qquad |E_{(2,2)}(D_{\mathrm{n}}^{(r)})| = r(\mathrm{n}-2),$$

$$E_{(2,2r)}(D_{\mathrm{n}}^{(r)}) = \{\, uv \in E(D_{\mathrm{n}}^{(r)}) \;:\; \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{2, 2r\} \,\}, \qquad |E_{(2,2r)}(D_{\mathrm{n}}^{(r)})| = 2r,$$

with

$$\mathcal{T}(2,2) = 2, \qquad \mathcal{T}(2,2r) = 2r.$$

Hence

$${}^{m}\mathrm{BI}_{\Psi}(D_{\mathrm{n}}^{(r)}) = r(\mathrm{n}-2)\,\frac{\Psi(2,2)}{2} + 2r\,\frac{\Psi(2,2r)}{2r} = \frac{r(\mathrm{n}-2)}{2}\,\Psi(2,2) + \Psi(2,2r).$$

ARTICLE IN PRESS

d-cube (hypercube) graph Q_d

The d-dimensional cube Q_d is d-regular with $|V(Q_d)| = 2^d$ and

$$|E(Q_d)| = d\,2^{d-1}, \qquad \phi_G(\mathrm{u}) = d \text{ for all } u \in V(Q_d).$$

Every edge lies in the single degree class $E_{(d,d)}(Q_d)$, and

$$\mathcal{T}(d,d) = d + d - 2 = 2(d-1).$$

Consequently, for any symmetric kernel Ψ and $d \geq 2$,

$${}^{m}\mathrm{BI}_{\Psi}(Q_d) = d\,2^{d-1}\,\frac{\Psi(d,d)}{2(d-1)} = \frac{2^{d-2}\,d}{d-1}\,\Psi(d,d),$$

which coincides with the regular-graph expression in Corollary 2.4.

3 Bond-Based Degree Indices and Notation

Let $G = (V, E)$ be a simple, finite, connected, undirected graph. For a vertex $u \in V$, we denote its degree by $\phi_G(\mathrm{u})$. In some formulas (following the original bond-based literature) the notation $\phi_G(\mathrm{u})$ also appears; *throughout this paper we use*

$$\phi_G(\mathrm{u}) \equiv \delta_G(\mathrm{u}) \equiv deg_G(\mathrm{u})$$

for the same vertex degree. This convention avoids ambiguity while keeping compatibility with earlier definitions.

For an edge $uv \in E(G)$ we introduce the shorthand

$$\mathcal{S}_{uv} = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}), \qquad \mathcal{P}_{uv} = \phi_G(\mathrm{u})\,\phi_G(\mathrm{v}),$$

and define its (edge/bond) degree by

$$\mathcal{T}(uv) \;:=\; \mathcal{S}_{uv} - 2 \;=\; \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}) - 2.$$

Thus $\mathcal{T}(uv)$ counts the number of edges incident with u or v other than uv itself; in particular, $\mathcal{T}(uv) = 0$ if and only if $\phi_G(\mathrm{u}) = \phi_G(\mathrm{v}) = 1$ (i.e., $G \cong K_2$ in the connected case). In particular, $\mathcal{T}(uv) > 0$ for every edge whenever G is connected and $|V(G)| \geq 3$.

3.1 Closed forms on benchmark graph families

In this subsection we record explicit closed-form expressions for the modified bond-based indices ${}^{m}\mathrm{BM}_1$, ${}^{m}\mathrm{BM}_2$, ${}^{m}\mathrm{BF}$, ${}^{m}\mathrm{BY}$, ${}^{m}\mathrm{BPC}$, ${}^{m}\mathrm{BSC}$, ${}^{m}\mathrm{BN}$, ${}^{m}\mathrm{BABC}$, ${}^{m}\mathrm{BCAB}$, ${}^{m}\mathrm{BGO}_1$, ${}^{m}\mathrm{BGO}_2$, ${}^{m}\mathrm{BGA}$, ${}^{m}\mathrm{BH}$, ${}^{m}\mathrm{BMP}$, ${}^{m}\mathrm{BS}$, ${}^{m}\mathrm{BD}$ for several standard graph families. All formulas follow directly from Proposition 2.6 by inserting the corresponding edge–degree partitions.

Theorem 3.1. *Let G be one of the standard graph families*

$$P_{\mathrm{n}}, \quad C_{\mathrm{n}}, \quad K_{\mathrm{n}}, \quad K_{m,\mathrm{n}}, \quad S_{\mathrm{n}} \cong K_{1,\mathrm{n}-1},$$

with $\mathrm{n} \geq 3$ and $m, \mathrm{n} \geq 2$ as appropriate. Then, for each of the modified bond-based indices the corresponding closed-form expressions are given in Table 1.

ARTICLE IN PRESS

Proof sketch. For each family $G \in \{P_{\mathrm{n}}, C_{\mathrm{n}}, K_{\mathrm{n}}, K_{m,\mathrm{n}}, S_{\mathrm{n}}\}$, we first determine the degree sequence and partition the edge set into classes

$$E_{(a,b)}(G) \;=\; \big\{\, uv \in E(G) : \{\phi_G(\mathrm{u}), \phi_G(\mathrm{v})\} = \{a, b\} \,\big\},$$

with cardinalities $|E_{(a,b)}(G)|$ as in the previous subsection.

For completeness, we note that the edge partitions $E_{(a,b)}(G)$ for these basic families are standard and follow immediately from their degree sequences; follow immediately from their degree sequences and can be verified directly from the definitions. . In particular, once the multiplicities $|E_{(a,b)}(G)|$ are fixed, the evaluation of ${}^{m}\mathrm{BI}_{\Psi}(G)$ reduces to a finite sum of constants and is routine.

By Proposition 2.6, any modified bond-based index associated with a symmetric kernel Ψ can be written as

$${}^{m}\mathrm{BI}_{\Psi}(G) = \sum_{(a,b)} \frac{\Psi(a,b)}{a+b-2}\, \big|E_{(a,b)}(G)\big|,$$

where the sum ranges over the degree pairs (a, b) that occur among the edges of G. Substituting the appropriate kernels Ψ for each index,

$$\Psi(x,y) \in \big\{x+y,\; xy,\; x^2+y^2,\; x^3+y^3,\; (xy)^{-1/2},\; (x+y)^{-1/2},\; \sqrt{x+y},\; \ldots\big\},$$

and the corresponding edge partitions of P_{n}, C_{n}, K_{n}, $K_{m,\mathrm{n}}$ and S_{n}, we obtain the explicit expressions appearing in Table 1 after straightforward algebraic simplification.

We illustrate the procedure on P_{n}; the remaining families are analogous, differing only in the list of occurring degree-pairs and their multiplicities. For instance, for P_{n} we have the edge classes

$$|E_{(1,2)}(P_{\mathrm{n}})| = 2, \qquad |E_{(2,2)}(P_{\mathrm{n}})| = \mathrm{n}-3, \qquad \mathcal{T}(1,2) = 1, \quad \mathcal{T}(2,2) = 2,$$

and hence

$${}^{m}\mathrm{BI}_{\Psi}(P_{\mathrm{n}}) = 2\,\Psi(1,2) + (\mathrm{n}-3)\,\frac{\Psi(2,2)}{2},$$

from which the entries in the row corresponding to P_{n} follow by choosing each kernel Ψ in turn. The remaining cases are treated analogously and give the formulas in the rows for C_{n}, K_{n}, $K_{m,\mathrm{n}}$ and S_{n}. Note that for $n \geq 3$ one has $\mathcal{T}(uv) = \phi(u) + \phi(v) - 2 \geq 1$ for every edge in each of these families (including S_n), hence all modified bond-based indices are well-defined.

□

Corollary 3.2. *The full family of modified bond-based indices introduced in Def. 1 for each of the benchmark graph families $G \in \{F_{\mathrm{n}}, W_{\mathrm{n}}, B_{\mathrm{n}}, D_{\mathrm{n}}^{(m)}, Q_d\}$ are given in the Table 2*

3.2 Bounds for modified bond-based indices

In this subsection we record simple degree–extreme (min–max) bounds for a representative subset of the modified bond-based indices. Throughout, let $G = (V, E)$ be a simple, connected graph without pendant edges (equivalently, with minimum degree $\delta(G) \geq 2$). Write $\mathrm{m} = |V|$, $m = |E|$, $\delta = \delta(G) \geq 2$ and $\Delta = \Delta(G)$ for the order, size, minimum and maximum degrees of G, respectively. For each edge $uv \in E$ set

$$\mathcal{S}_{uv} = \phi_G(\mathrm{u}) + \phi_G(\mathrm{v}), \qquad \mathcal{P}_{uv} = \phi_G(\mathrm{u})\,\phi_G(\mathrm{v}), \qquad \mathcal{T}(uv) = \mathcal{S}_{uv} - 2,$$

ARTICLE IN PRESS

Index	P_{n} : $\mathrm{n} \geq 3$	C_{n} : $\mathrm{n} \geq 3$	K_{n} : $\mathrm{n} \geq 3$	$K_{m,\mathrm{n}}$: $m, \mathrm{n} \geq 2$	$S_{\mathrm{n}} \cong K_{1,\mathrm{n}-1}$: $\mathrm{n} \geq 3$
${}^{m}\mathrm{BM}_1$	$2\mathrm{n}$	$2\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^2}{2\,(\mathrm{n}-2)}$	$\frac{mn(m+\mathrm{n})}{m+\mathrm{n}-2}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)}{\mathrm{n}-2}$
${}^{m}\mathrm{BM}_2$	$2\mathrm{n}-2$	$2\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^3}{4\,(\mathrm{n}-2)}$	$\frac{m^2\mathrm{n}^2}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)^2}{\mathrm{n}-2}$
${}^{m}\mathrm{BF}$	$4\mathrm{n}-2$	$4\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^3}{2\,(\mathrm{n}-2)}$	$\frac{mn(m^2+\mathrm{n}^2)}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)\big((\mathrm{n}-1)^2+1\big)}{\mathrm{n}-2}$
${}^{m}\mathrm{BY}$	$8\mathrm{n}-6$	$8\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^4}{2\,(\mathrm{n}-2)}$	$\frac{mn(m^3+\mathrm{n}^3)}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)\big((\mathrm{n}-1)^3+1\big)}{\mathrm{n}-2}$
${}^{m}\mathrm{BPC}$	$\sqrt{2}+\frac{\mathrm{n}-3}{4}$	$\frac{\mathrm{n}}{4}$	$\frac{\mathrm{n}}{4\,(\mathrm{n}-2)}$	$\frac{\sqrt{mn}}{m+\mathrm{n}-2}$	$\frac{\sqrt{\mathrm{n}-1}}{\mathrm{n}-2}$
${}^{m}\mathrm{BSC}$	$\frac{2\sqrt{3}}{3}+\frac{\mathrm{n}-3}{4}$	$\frac{\mathrm{n}}{4}$	$\frac{\mathrm{n}\sqrt{2\mathrm{n}-2}}{8\,(\mathrm{n}-2)}$	$\frac{mn}{\sqrt{m+\mathrm{n}}\,(m+\mathrm{n}-2)}$	$\frac{\mathrm{n}-1}{\sqrt{\mathrm{n}}\,(\mathrm{n}-2)}$
${}^{m}\mathrm{BN}$	$2\sqrt{3}+\mathrm{n}-3$	n	$\frac{\sqrt{2}\,\mathrm{n}\,(\mathrm{n}-1)^{3/2}}{4\,(\mathrm{n}-2)}$	$\frac{mn\sqrt{m+\mathrm{n}}}{m+\mathrm{n}-2}$	$\frac{\sqrt{\mathrm{n}}\,(\mathrm{n}-1)}{\mathrm{n}-2}$
${}^{m}\mathrm{BABC}$	$\frac{\sqrt{2}\,(\mathrm{n}+1)}{4}$	$\frac{\sqrt{2}}{4}\,\mathrm{n}$	$\frac{\sqrt{2}\,\mathrm{n}}{4\sqrt{\mathrm{n}-2}}$	$\frac{\sqrt{mn}}{\sqrt{m+\mathrm{n}-2}}$	$\frac{\sqrt{\mathrm{n}-1}}{\sqrt{\mathrm{n}-2}}$
${}^{m}\mathrm{BCAB}$	$\frac{\sqrt{2}\,(\mathrm{n}+1)}{2}$	$\frac{\sqrt{2}}{2}\,\mathrm{n}$	$\frac{\sqrt{2}\,\mathrm{n}\,(\mathrm{n}-1)^2}{8\,(\mathrm{n}-2)^{3/2}}$	$\frac{m^{3/2}\mathrm{n}^{3/2}}{(m+\mathrm{n}-2)^{3/2}}$	$\frac{(\mathrm{n}-1)^{3/2}}{(\mathrm{n}-2)^{3/2}}$
${}^{m}\mathrm{BGO}_1$	$4\mathrm{n}-4$	$4\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^3}{2\,(\mathrm{n}-2)}$	$\frac{2m^2\mathrm{n}^2}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)(2\mathrm{n}-2)}{\mathrm{n}-2}$
${}^{m}\mathrm{BGO}_2$	$8\mathrm{n}-12$	$8\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^4}{2\,(\mathrm{n}-2)}$	$\frac{m^2\mathrm{n}^2(m+\mathrm{n})}{m+\mathrm{n}-2}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^2}{\mathrm{n}-2}$
${}^{m}\mathrm{BGA}$	$\frac{4\sqrt{2}}{3}+\frac{\mathrm{n}-3}{2}$	$\frac{\mathrm{n}}{2}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)}{4\,(\mathrm{n}-2)}$	$\frac{2m^{3/2}\mathrm{n}^{3/2}}{(m+\mathrm{n})(m+\mathrm{n}-2)}$	$\frac{2\,(\mathrm{n}-1)^{3/2}}{\mathrm{n}\,(\mathrm{n}-2)}$
${}^{m}\mathrm{BH}$	$\frac{4}{3}+\frac{\mathrm{n}-3}{4}$	$\frac{\mathrm{n}}{4}$	$\frac{\mathrm{n}}{4\,(\mathrm{n}-2)}$	$\frac{2mn}{(m+\mathrm{n})(m+\mathrm{n}-2)}$	$\frac{2\,(\mathrm{n}-1)}{\mathrm{n}\,(\mathrm{n}-2)}$
${}^{m}\mathrm{BMP}$	$2+\sqrt{2}\,(\mathrm{n}-1)$	$\sqrt{2}\,\mathrm{n}$	$\frac{\mathrm{n}\,(\mathrm{n}-1)^{3/2}}{2\,(\mathrm{n}-2)}$	$\frac{mn(\sqrt{m}+\sqrt{\mathrm{n}})}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)\big(\sqrt{\mathrm{n}-1}+1\big)}{\mathrm{n}-2}$
${}^{m}\mathrm{BS}$	$2\sqrt{5}+\sqrt{2}\,(\mathrm{n}-3)$	$\sqrt{2}\,\mathrm{n}$	$\frac{\sqrt{2}\,\mathrm{n}\,(\mathrm{n}-1)^2}{4\,(\mathrm{n}-2)}$	$\frac{mn\sqrt{m^2+\mathrm{n}^2}}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)\sqrt{(\mathrm{n}-1)^2+1}}{\mathrm{n}-2}$
${}^{m}\mathrm{BD}$	$2\mathrm{n}$	$2\mathrm{n}$	$\frac{\sqrt{2}\,\mathrm{n}\,(\mathrm{n}-1)^{5/2}}{4\,(\mathrm{n}-2)}$	$\frac{mn\sqrt{m^3+\mathrm{n}^3}}{m+\mathrm{n}-2}$	$\frac{(\mathrm{n}-1)\sqrt{(\mathrm{n}-1)^3+1}}{\mathrm{n}-2}$

Table 1: Closed-form expressions of modified bond-based indices for paths P_{n}, cycles C_{n}, complete graphs K_{n}, complete bipartite graphs $K_{m,\mathrm{n}}$, and stars $S_{\mathrm{n}} \cong K_{1,\mathrm{n}-1}$.

ARTICLE IN PRESS

so that

$$2\delta \leq \mathcal{S}_{uv} \leq 2\Delta, \qquad \delta^2 \leq \mathcal{P}_{uv} \leq \Delta^2, \qquad 2\delta - 2 \leq \mathcal{T}(uv) \leq 2\Delta - 2 \quad (uv \in E).$$

ARTICLE IN PRESS

Table 2: Closed forms of all modified bond-based indices for the friendship graph F_{n}, wheel W_{n}, book B_{n}, Dutch windmill $D_{\mathrm{n}}^{(r)}$ and d-cube Q_d.

Index	F_{n}, $\mathrm{n}\geq 3$	W_{n}, $\mathrm{n}\geq 3$	B_{n}, $\mathrm{n}\geq 3$	$D_{\mathrm{n}}^{(r)}$, $\mathrm{n}\geq 3$, $r\geq 2$	Q_d, $d\geq 2$
${}^{m}\mathrm{BM}_1$	$4\mathrm{n}+2$	$\frac{(\mathrm{n}-1)(5\mathrm{n}+9)}{2(\mathrm{n}+1)}$	$\frac{4\mathrm{n}^3-7\mathrm{n}^2-2\mathrm{n}+1}{(\mathrm{n}-1)(\mathrm{n}-2)}$	$2(r\mathrm{n}-r+1)$	$\frac{2^{d-1}d^2}{d-1}$
${}^{m}\mathrm{BM}_2$	$6\mathrm{n}$	$\frac{3(\mathrm{n}-1)(7\mathrm{n}+3)}{4(\mathrm{n}+1)}$	$\frac{13\mathrm{n}^2-26\mathrm{n}+1}{2(\mathrm{n}-2)}$	$2r\mathrm{n}$	$\frac{2^{d-2}d^3}{d-1}$
${}^{m}\mathrm{BF}$	$4\mathrm{n}^2+4\mathrm{n}+4$	$\frac{(\mathrm{n}-1)\left(2\mathrm{n}^2+9\mathrm{n}+27\right)}{2(\mathrm{n}+1)}$	$\frac{2\mathrm{n}^4-3\mathrm{n}^3+3\mathrm{n}^2-9\mathrm{n}-1}{(\mathrm{n}-1)(\mathrm{n}-2)}$	$4\left(r^2+r\mathrm{n}-2r+1\right)$	$\frac{2^{d-1}d^3}{d-1}$
${}^{m}\mathrm{BY}$	$8\mathrm{n}^3+8\mathrm{n}+8$	$\frac{(\mathrm{n}-1)\left(2\mathrm{n}^3+27\mathrm{n}+81\right)}{2(\mathrm{n}+1)}$	$\frac{2\mathrm{n}^5-9\mathrm{n}^4+22\mathrm{n}^3-16\mathrm{n}^2-16\mathrm{n}+1}{(\mathrm{n}-1)(\mathrm{n}-2)}$	$8\left(r^3+r\mathrm{n}-2r+1\right)$	$\frac{2^{d-1}d^4}{d-1}$
${}^{m}\mathrm{BPC}$	$\frac{\mathrm{n}}{4}+\frac{1}{2\sqrt{\mathrm{n}}}$	$(\mathrm{n}-1)\left(\frac{1}{12}+\frac{1}{(\mathrm{n}+1)\sqrt{3\mathrm{n}}}\right)$	$\frac{\mathrm{n}}{4}+\frac{2\mathrm{n}}{(\mathrm{n}-1)\sqrt{2\mathrm{n}-2}}+\frac{1}{(2\mathrm{n}-4)(\mathrm{n}-1)}$	$\frac{r(\mathrm{n}-2)}{4}+\frac{\sqrt{r}}{2\sqrt{2}}$	$\frac{2^{d-2}}{d-1}$
${}^{m}\mathrm{BSC}$	$\frac{\mathrm{n}}{4}+\frac{\sqrt{2}}{2\sqrt{\mathrm{n}+1}}$	$(\mathrm{n}-1)\left(\frac{1}{4\sqrt{6}}+\frac{1}{(\mathrm{n}+1)\sqrt{\mathrm{n}+3}}\right)$	$\frac{\mathrm{n}}{4}+\frac{2\mathrm{n}}{(\mathrm{n}-1)\sqrt{\mathrm{n}+1}}+\frac{1}{(2\mathrm{n}-4)\sqrt{2\mathrm{n}-2}}$	$\frac{r(\mathrm{n}-2)}{4}+\frac{\sqrt{r}}{2\sqrt{r+1}}$	$\frac{2^{d-2}\sqrt{d}}{\sqrt{2}\,(d-1)}$
${}^{m}\mathrm{BN}$	$n+\sqrt{2\mathrm{n}+2}$	$(\mathrm{n}-1)\left(\frac{\sqrt{6}}{4}+\frac{\sqrt{\mathrm{n}+3}}{\mathrm{n}+1}\right)$	$n+\frac{2\mathrm{n}\sqrt{\mathrm{n}+1}}{\mathrm{n}-1}+\frac{\sqrt{2\mathrm{n}-2}}{2\mathrm{n}-4}$	$\frac{r(\mathrm{n}+2)}{2}+\sqrt{2r+2}$	$\frac{2^{d-2}d\sqrt{2d}}{d-1}$
${}^{m}\mathrm{BABC}$	$\frac{\sqrt{2}(\mathrm{n}+2)}{4}$	$(\mathrm{n}-1)\left(\frac{1}{6}+\frac{1}{\sqrt{3\mathrm{n}(\mathrm{n}+1)}}\right)$	$\frac{\sqrt{2}\,\mathrm{n}}{4}+\frac{\sqrt{2}\,\mathrm{n}}{\mathrm{n}-1}+\frac{1}{(\mathrm{n}-1)\sqrt{2(\mathrm{n}-2)}}$	$\frac{r(\mathrm{n}-2)}{4}+\frac{1}{2\sqrt{2r}}$	$\frac{2^{d-2}\sqrt{2d-2}}{d-1}$
${}^{m}\mathrm{BCAB}$	$\frac{\sqrt{2}(\mathrm{n}+2)}{2}$	$(\mathrm{n}-1)\left(\frac{3}{8}+\frac{\sqrt{3\mathrm{n}}}{(\mathrm{n}+1)^{3/2}}\right)$	$\frac{\mathrm{n}}{\sqrt{2}}+\frac{2\mathrm{n}\sqrt{2}}{\mathrm{n}-1}+\frac{(\mathrm{n}-1)}{2\sqrt{2}\,(\mathrm{n}-2)^{3/2}}$	$\sqrt{2r}+\frac{r\mathrm{n}}{\sqrt{2}}$	$\frac{2^{d-2}d^2}{(d-1)\sqrt{2d-2}}$
${}^{m}\mathrm{BGO}_1$	$12\mathrm{n}$	$(\mathrm{n}-1)\left(\frac{9}{2}+\frac{6\mathrm{n}}{\mathrm{n}+1}\right)$	$4\mathrm{n}+\frac{2\mathrm{n}(4\mathrm{n}-4)}{\mathrm{n}-1}+\frac{(\mathrm{n}-1)(2\mathrm{n}-2)}{2\mathrm{n}-4}$	$4r\mathrm{n}$	$\frac{2^d d^3}{d-1}$
${}^{m}\mathrm{BGO}_2$	$8\mathrm{n}(\mathrm{n}+2)$	$(\mathrm{n}-1)\left(\frac{27}{2}+\frac{3\mathrm{n}(\mathrm{n}+3)}{\mathrm{n}+1}\right)$	$8\mathrm{n}+\frac{2\mathrm{n}\left(4\mathrm{n}+2(\mathrm{n}-1)^2-4\right)}{\mathrm{n}-1}+\frac{2(\mathrm{n}-1)^3}{2\mathrm{n}-4}$	$8r\mathrm{n}^2$	$\frac{2^d d^4}{d-1}$
${}^{m}\mathrm{BGA}$	$\frac{4\sqrt{\mathrm{n}}+n(\mathrm{n}+1)}{2(\mathrm{n}+1)}$	$(\mathrm{n}-1)\left(\frac{1}{4}+\frac{2\sqrt{3\mathrm{n}}}{(\mathrm{n}+1)(\mathrm{n}+3)}\right)$	$\frac{\mathrm{n}}{2}+\frac{4\mathrm{n}\sqrt{2(\mathrm{n}-1)}}{(\mathrm{n}+1)(\mathrm{n}-1)}+\frac{1}{2(\mathrm{n}-2)}$	$\frac{r\left(r(\mathrm{n}-2)+2\mathrm{n}\right)}{r+1}$	$\frac{2^{d-2}d}{d-1}$
${}^{m}\mathrm{BH}$	$\frac{n(\mathrm{n}+1)+4}{4(\mathrm{n}+1)}$	$(\mathrm{n}-1)\left(\frac{1}{12}+\frac{2}{(\mathrm{n}+1)(\mathrm{n}+3)}\right)$	$\frac{\mathrm{n}}{4}+\frac{4\mathrm{n}}{(\mathrm{n}-1)(\mathrm{n}+1)}+\frac{1}{(2\mathrm{n}-2)(\mathrm{n}-2)}$	$\frac{r(\mathrm{n}+2)}{r+1}$	$\frac{2^{d-2}}{d-1}$
${}^{m}\mathrm{BMP}$	$\sqrt{2}\,(\sqrt{\mathrm{n}}+\mathrm{n}+1)$	$(\mathrm{n}-1)\left(\frac{\sqrt{3}}{2}+\frac{\sqrt{3}+\sqrt{\mathrm{n}}}{\mathrm{n}+1}\right)$	$n\sqrt{2}+\frac{2\mathrm{n}(\sqrt{2}+\sqrt{\mathrm{n}-1})}{\mathrm{n}-1}+\frac{\sqrt{\mathrm{n}-1}}{\mathrm{n}-2}$	$2\sqrt{2r}+r\sqrt{\mathrm{n}}$	$\frac{2^{d-1}d\sqrt{d}}{d-1}$
${}^{m}\mathrm{BS}$	$\sqrt{2}\,\mathrm{n}+2\sqrt{\mathrm{n}^2+1}$	$\frac{3\sqrt{2}}{4}(\mathrm{n}-1)+\frac{(\mathrm{n}-1)\sqrt{\mathrm{n}^2+9}}{\mathrm{n}+1}$	$\sqrt{2}\,\mathrm{n}+\frac{2\mathrm{n}\sqrt{(\mathrm{n}-1)^2+4}}{\mathrm{n}-1}+\frac{(\mathrm{n}-1)\sqrt{2}}{2(\mathrm{n}-2)}$	$\sqrt{2}\,r(\mathrm{n}-2)+2\sqrt{1+r^2}$	$\frac{2^{d-2}d^2\sqrt{2}}{d-1}$
${}^{m}\mathrm{BD}$	$2\mathrm{n}+2\sqrt{2\mathrm{n}^3+2}$	$\frac{3\sqrt{6}}{4}(\mathrm{n}-1)+\frac{(\mathrm{n}-1)\sqrt{\mathrm{n}^3+27}}{\mathrm{n}+1}$	$2\mathrm{n}+\frac{2\mathrm{n}\sqrt{8+(\mathrm{n}-1)^3}}{\mathrm{n}-1}+\frac{\sqrt{2(\mathrm{n}-1)^3}}{2(\mathrm{n}-2)}$	$4r(\mathrm{n}-2)+\sqrt{8+8r^3}$	$\frac{2^{d-2}d^{5/2}\sqrt{2}}{d-1}$

ARTICLE IN PRESS

ARTICLE IN PRESS

Theorem 3.3. *Under the above assumptions, the following bounds hold:*

$$
\begin{aligned}
m\,\frac{2\Delta}{2\Delta-2} \;&\le\; {}^{m}\mathrm{BM}_1(G) \;\le\; m\,\frac{2\delta}{2\delta-2},\\
m\,\frac{\delta^2}{2\Delta-2} \;&\le\; {}^{m}\mathrm{BM}_2(G) \;\le\; m\,\frac{\Delta^2}{2\delta-2},\\
m\,\frac{2\delta^2}{2\Delta-2} \;&\le\; {}^{m}\mathrm{BF}(G) \;\le\; m\,\frac{2\Delta^2}{2\delta-2},\\
m\,\frac{2\delta^3}{2\Delta-2} \;&\le\; {}^{m}\mathrm{BY}(G) \;\le\; m\,\frac{2\Delta^3}{2\delta-2},\\
m\,\frac{1}{(2\Delta-2)\,\Delta} \;&\le\; {}^{m}\mathrm{BPC}(G) \;\le\; m\,\frac{1}{(2\delta-2)\,\delta},\\
m\,\frac{1}{(2\Delta-2)\sqrt{2\Delta}} \;&\le\; {}^{m}\mathrm{BSC}(G) \;\le\; m\,\frac{1}{(2\delta-2)\sqrt{2\delta}},\\
m\,\frac{\sqrt{2\Delta}}{2\Delta-2} \;&\le\; {}^{m}\mathrm{BN}(G) \;\le\; m\,\frac{\sqrt{2\delta}}{2\delta-2},\\
m\,\frac{1}{\sqrt{\Delta^2(2\Delta-2)}} \;&\le\; {}^{m}\mathrm{BABC}(G) \;\le\; m\,\frac{1}{\sqrt{\delta^2(2\delta-2)}},\\
m\,\frac{\delta}{(2\Delta-2)^{3/2}} \;&\le\; {}^{m}\mathrm{BCAB}(G) \;\le\; m\,\frac{\Delta}{(2\delta-2)^{3/2}}.
\end{aligned}
$$

All bounds are sharp: simultaneous equality in the lower and upper bounds for any of the above indices holds if and only if G is regular (i.e., $\delta = \Delta$).

Proof sketch. Each inequality is obtained by bounding the edgewise kernel using the extremal values of $\mathcal{S}_{uv}$, $\mathcal{P}_{uv}$ and $\mathcal{T}(uv)$. For example,

$$
{}^{m}\mathrm{BM}_1(G) = \sum_{uv\in E(G)} \frac{\mathcal{S}_{uv}}{\mathcal{T}(uv)} = \sum_{uv\in E(G)} f(\mathcal{S}_{uv}), \qquad f(S) = \frac{S}{S-2},
$$

Indeed, $f'(S) = \frac{-2}{(S-2)^2} < 0$ for $S > 2$, so f is strictly decreasing on $(2,\infty)$. Since $\delta \ge 2$ in this section, we have $\mathcal{S}_{uv} \ge 2\delta \ge 4$, hence $f(\mathcal{S}_{uv})$ is well-defined for every edge. Since $2\delta \le \mathcal{S}_{uv} \le 2\Delta$, we obtain $f(2\Delta) \le f(\mathcal{S}_{uv}) \le f(2\delta)$ for every edge, which yields the bounds for ${}^{m}\mathrm{BM}_1(G)$ after summation.

For ${}^{m}\mathrm{BM}_2(G) = \sum \mathcal{P}_{uv}/\mathcal{T}(uv)$ we use $\delta^2 \le \mathcal{P}_{uv} \le \Delta^2$ and $2\delta - 2 \le \mathcal{T}(uv) \le 2\Delta - 2$, Since $\mathcal{P}_{uv} > 0$ and $\mathcal{T}(uv) > 0$, we may bound $\frac{\mathcal{P}_{uv}}{\mathcal{T}(uv)}$ by combining the extremal bounds on numerator and denominator (monotonicity of $x \mapsto x$ and $x \mapsto 1/x$ on $(0,\infty)$): together with the monotonicity of $x \mapsto x$ and $x \mapsto 1/x$ on $(0,\infty)$, to enclose each term between $\delta^2/(2\Delta-2)$ and $\Delta^2/(2\delta-2)$. The bounds for ${}^{m}\mathrm{BF}$ and ${}^{m}\mathrm{BY}$ follow analogously from $2\delta^2 \le \phi_G(\mathrm{u})^2 + \phi_G(\mathrm{v})^2 \le 2\Delta^2$ and $2\delta^3 \le \phi_G(\mathrm{u})^3 + \phi_G(\mathrm{v})^3 \le 2\Delta^3$.

For the connectivity–type indices, we rewrite each summand in terms of $\mathcal{S}_{uv}$, $\mathcal{P}_{uv}$ and $\mathcal{T}(uv)$,

ARTICLE IN PRESS

for instance

$$\frac{1}{\mathcal{T}(uv)\sqrt{\mathcal{P}_{uv}}} = \frac{1}{\mathcal{T}(uv)} \cdot \frac{1}{\sqrt{\mathcal{P}_{uv}}}, \qquad \frac{\sqrt{\mathcal{S}_{uv}}}{\mathcal{T}(uv)} = \frac{\sqrt{\mathcal{S}_{uv}}}{\mathcal{S}_{uv} - 2}, \qquad \frac{1}{\sqrt{\mathcal{P}_{uv}\mathcal{T}(uv)}} = \frac{1}{\sqrt{\mathcal{P}_{uv}}} \cdot \frac{1}{\sqrt{\mathcal{T}(uv)}},$$

and then use

$$\delta \le \sqrt{\mathcal{P}_{uv}} \le \Delta, \quad \sqrt{2\delta} \le \sqrt{\mathcal{S}_{uv}} \le \sqrt{2\Delta}, \quad \frac{1}{2\Delta - 2} \le \frac{1}{\mathcal{T}(uv)} \le \frac{1}{2\delta - 2},$$

together with the fact that the functions

$$S \mapsto \frac{\sqrt{S}}{S-2}, \qquad S \mapsto \frac{1}{(S-2)\sqrt{S}}$$

are decreasing on $[2\delta, \infty)$. (For example, differentiating shows both derivatives are negative for $S > 2$.) Summing the resulting per–edge inequalities over $E(G)$ yields the stated bounds. Sharpness is attained by any regular graph (e.g., C_n or K_n), for which $\mathcal{S}_{uv}$, $\mathcal{P}_{uv}$ and $\mathcal{T}(uv)$ are constant over $E(G)$; hence each per-edge bound is an equality and so is the summed bound. Conversely, if equality holds in the upper (or lower) bound, then equality must hold in the corresponding per-edge estimate for every edge, forcing $\mathcal{S}_{uv}$ (and thus the endpoint degrees) to be constant on all edges. By connectivity this implies all vertices have the same degree, i.e., G is regular and $\delta = \Delta$.

□

Remark 3.4. *Analogous degree–extreme estimates can be derived for the remaining modified bond-based indices (such as* ${}^{m}\mathrm{BGO}_1$, ${}^{m}\mathrm{BGO}_2$, ${}^{m}\mathrm{BGA}$, ${}^{m}\mathrm{BH}$, ${}^{m}\mathrm{BMP}$, ${}^{m}\mathrm{BS}$ *and* ${}^{m}\mathrm{BD}$*).*

Corollary 3.5. *Let G be a k-regular graph with $k \ge 2$ and $m = |E(G)|$. Then all the degree–extreme bounds in Theorem 3.3 collapse to equalities, and the corresponding modified bond-based indices admit the closed forms*

$$
\begin{aligned}
{}^{m}\mathrm{BM}_1(G) &= m\,\frac{2k}{2k-2} = m\,\frac{k}{k-1}, & {}^{m}\mathrm{BM}_2(G) &= m\,\frac{k^2}{2k-2} = m\,\frac{k^2}{2(k-1)},\\
{}^{m}\mathrm{BF}(G) &= m\,\frac{2k^2}{2k-2} = m\,\frac{k^2}{k-1}, & {}^{m}\mathrm{BY}(G) &= m\,\frac{2k^3}{2k-2} = m\,\frac{k^3}{k-1},\\
{}^{m}\mathrm{BPC}(G) &= m\,\frac{1}{(2k-2)\,k}, & {}^{m}\mathrm{BSC}(G) &= m\,\frac{1}{(2k-2)\sqrt{2k}},\\
{}^{m}\mathrm{BN}(G) &= m\,\frac{\sqrt{2k}}{2k-2}, & {}^{m}\mathrm{BABC}(G) &= m\,\frac{1}{\sqrt{k^2(2k-2)}} = m\,\frac{1}{k\sqrt{2k-2}},\\
{}^{m}\mathrm{BCAB}(G) &= m\,\frac{k}{(2k-2)^{3/2}}.
\end{aligned}
$$

In particular, for k-regular graphs each of these modified bond-based indices is a simple scalar multiple of the size m, with the scalar depending only on k.

ARTICLE IN PRESS

QSAR/QSPR Modelling and Benchmark Evaluation

4 Methods

4.1 Overall modelling strategy

On the practical side, this study develops a systematic, multi-task QSAR/QSPR framework for small-molecule antibacterial agents. The central question is whether a recently proposed family of modified bond-based indices (^{m}BI) descriptors can enhance structure–property modelling, either on their own or when combined with standard physicochemical descriptors. To address this, we consider a panel of ten continuous targets that jointly characterise molecular physicochemical behaviour, and we benchmark a large and heterogeneous set of regression algorithms under three descriptor scenarios, using a common train–test split and harmonised evaluation metrics.

The modelling pipeline comprises four stages: (i) curation of a consistent dataset of small molecules and selection of ten continuous target properties; (ii) computation of modified bond-based indices descriptors and extraction of physicochemical descriptors, followed by definition of three feature scenarios; (iii) multi-task training of a broad "model zoo" of regression algorithms with three-fold cross-validation and external test evaluation; and (iv) aggregation of per-task results into macro-averaged and global metrics to compare descriptor scenarios at the multi-target level.

4.2 Dataset and target properties

All compounds were obtained from the ChEMBL database, a manually curated, open and FAIR resource of bioactive molecules with drug-like properties [88]. For each record, ChEMBL provides structural identifiers such as SMILES strings, Molecule ChEMBL IDs and IUPAC InChIKeys, together with an extensive set of physicochemical descriptors and ligand-based drug-likeness metrics. Antibacterial activity measurements, including minimum inhibitory concentration (MIC), are also available for many small-molecule agents. Since a single chemical structure can appear multiple times in ChEMBL (e.g., linked to different assays, targets or experimental settings), structural redundancy was removed by adopting a stringent uniqueness criterion based on the pair (Molecule ChEMBL ID, IUPAC InChIKey). All records sharing the same identifier pair were collapsed into a single molecular entry. Within this de-duplicated set, only molecules with complete descriptor fields and a quantitative antibacterial activity readout were retained. When multiple quantitative measurements were available for the same compound under comparable conditions, clearly inconsistent or non-quantitative entries were discarded and the remaining records were harmonised to a single standard value according to assay metadata.

From this curated collection we formed a panel of ten continuous targets, treated on an equal footing throughout the modelling and evaluation. Nine targets are taken directly from ChEMBL-provided physicochemical fields (e.g., size-, polarity-, hydrogen-bonding-, flexibility- and drug-likeness-related properties), and the remaining target represents antibacterial potency via a log-transformed activity measure. For a molecule i with a strictly positive standardised MIC value $\mathrm{MIC}_i > 0$, the transformed activity is defined as

$$\mathrm{pMIC}_i = -\log_{10}(\mathrm{MIC}_i), \tag{3}$$

which reduces skewness and aligns the activity scale with common QSAR/QSPR practice. After applying all filters and requiring completeness across targets and descriptor fields, the final analysis-

ARTICLE IN PRESS

ready dataset comprised $N = 3{,}219$ unique small-molecule entries. All subsequent modelling is performed on this fixed compound set to ensure strict comparability across targets, descriptor scenarios and regression algorithms.

Preprocessing note. All molecular graphs in the benchmark satisfy $|V(G)| \geq 3$, so $\mathcal{T}(uv) \geq 1$ for every bond uv and ^{m}BI descriptors are well-defined without any special handling of pendant edges. For completeness, if a dataset were to contain the degenerate connected case $G \cong K_2$ (the only situation where $\mathcal{T}(uv) = 0$), such entries can be trivially filtered or treated separately.

4.3 Descriptor construction and feature scenarios

Each molecule is represented by two complementary descriptor families. The first family consists of 16 modified bond-based indices (^{m}BI) descriptors, as seen in Table 3 ,defined as graph- and topology-based invariants on the molecular graph. For molecule i, this vector is denoted

$$\mathbf{x}_i^{({}^{m}\mathrm{BI})} \in \mathbb{R}^{16}. \tag{4}$$

These descriptors encode structural information through bond-centred weighting schemes and connectivity patterns and provide a compact structural representation independent of conventional physicochemical computations.

The second family comprises 10 physicochemical descriptors extracted from ChEMBL property fields (seen in Table 4). For molecule i,

$$\mathbf{x}_i^{(\mathrm{Physchem})} \in \mathbb{R}^{10}. \tag{5}$$

These descriptors summarise global physicochemical behaviour and are widely used in tabular QSAR/QSPR modelling.

ARTICLE IN PRESS

Computational complexity. For a molecular graph $G = (V, E)$, each ^{m}BI descriptor is an edge-sum whose summand depends only on endpoint degrees and the bond factor $\mathcal{T}(uv)$. After computing vertex degrees in $O(|E|)$ time, each descriptor is evaluated in a single pass over $E(G)$, hence $O(|E|)$ time and $O(|V|)$ memory. Computing the full set of 16 ^{m}BI descriptors is therefore $O(16|E|) = O(|E|)$, matching the scalability of standard degree-based and connectivity indices, and is computationally negligible compared with the downstream model-zoo training for datasets of this size.

Abbreviations: MW = Molecular Weight (g/mol); AlogP = calculated logP; PSA = Polar Surface Area (Å^2); HBA = Hydrogen Bond Acceptors; HBD = Hydrogen Bond Donors; RB = number of Rotatable Bonds; QED = QED weighted score; HA = number of Heavy Atoms; NP = NP-likeness score; MIC = MIC standard value (e.g., μg/mL).

To quantify the individual and complementary contributions of these descriptor families, three feature scenarios are defined and applied uniformly across all targets. In the *Modified bond-based indices descriptors (*^{m}BI*) scenario*, only $\mathbf{x}_i^{(\mathrm{mBI})}$ is used as the predictor vector. In the *Physicochemical descriptors scenario*, only $\mathbf{x}_i^{(\mathrm{Physchem})}$ is used. In the *Combined descriptors scenario*, the two families are concatenated to form a 27-dimensional hybrid representation,

$$\mathbf{x}_i^{(\mathrm{Comb})} = \left[\mathbf{x}_i^{(\mathrm{mBI})}, \mathbf{x}_i^{(\mathrm{Physchem})}\right] \in \mathbb{R}^{26}. \tag{6}$$

ARTICLE IN PRESS

Table 3: Selected compounds from the dataset, showing their molecular formulas and the corresponding values of the modified bond-based indices ${}^{m}\mathrm{BM}_1, {}^{m}\mathrm{BM}_2, {}^{m}\mathrm{BF}, \ldots, {}^{m}\mathrm{BD}$.

ID	Molecular Formula	${}^{m}\mathrm{BM}_1$	${}^{m}\mathrm{BM}_2$	${}^{m}\mathrm{BF}$	$\cdots$	${}^{m}\mathrm{BD}$
C1	$C_{28}H_{28}F_2N_4O_5$	77.17	88.75	192.83	$\cdots$	89.26
C2	$C_8H_{10}N_2S$	20.83	21.25	49.17	$\cdots$	23.99
C3	$C_{21}H_{29}N_2NaO_9S$	61.33	68.00	158.67	$\cdots$	74.28
C4	$C_{22}H_{24}N_2O_8$	59.20	69.03	164.33	$\cdots$	74.85
C5	$C_8H_6INO_2$	22.17	22.75	52.83	$\cdots$	25.66
C6	$C_{22}H_{24}N_6$	55.33	62.00	132.67	$\cdots$	62.25
C7	$C_{28}H_{44}N_4O_4$	68.23	73.65	169.17	$\cdots$	79.94
C8	$C_8H_{13}N_3O_4S$	28.83	29.92	75.83	$\cdots$	35.98
C9	$C_{17}H_{14}BrNO_4$	45.00	52.50	117.00	$\cdots$	53.70
C10	$C_{26}H_{23}F_3N_4O_6S$	73.90	82.65	196.50	$\cdots$	91.17
C11	$C_{27}H_{23}N_3O_5$	66.47	71.97	165.00	$\cdots$	77.97
C12	$C_{22}H_{29}N_7O_5$	66.17	73.75	164.83	$\cdots$	77.32
C13	$C_{17}H_{19}F_2N_3O_3$	48.17	53.75	122.83	$\cdots$	57.48
C14	$C_{17}H_{18}FN_3O_3$	46.67	54.00	119.33	$\cdots$	55.00
C15	$C_{16}H_{12}N_2O_2$	40.33	46.50	97.67	$\cdots$	45.28
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\ddots$	$\vdots$
C3218	$C_{17}H_{10}N_4O_3S_2$	50.67	57.50	126.33	$\cdots$	58.84
C3219	$C_{36}H_{74}N_{10}O_{12}$	110.00	121.00	278.00	$\cdots$	130.96

Table 4: Summary of basic physicochemical and QSPR-relevant properties (MW, AlogP, PSA, HBA, HBD, RB, QED, HA, NP-likeness and MIC) for the reference compound set.

ID	MW	AlogP	PSA	HBA	HBD	RB	QED	HA	NP	MIC
C1	538.55	3.50	105.39	8	2	9	0.40	39	-0.19	10.00
C2	166.25	1.28	38.91	2	1	2	0.67	11	-1.28	10.00
C3	508.53	0.91	151.70	9	3	9	0.24	33	0.73	25.00
C4	444.44	-0.35	181.62	9	6	2	0.33	32	1.81	3.13
C5	275.05	2.54	43.14	2	0	2	0.47	12	-0.66	4.80
C6	372.48	4.63	67.66	6	2	7	0.49	28	-0.86	20.00
C7	500.68	3.34	101.98	5	3	11	0.25	36	-0.16	31.00
C8	247.28	0.53	95.10	6	0	5	0.56	16	-2.05	0.80
C9	376.21	3.66	59.00	4	1	1	0.83	23	0.20	20.00
C10	576.55	2.04	101.45	6	2	3	0.69	29	-1.05	1.25
C11	469.50	1.89	144.91	6	6	5	0.15	35	0.01	5.00
C12	471.52	-0.79	160.88	11	4	8	0.33	34	0.41	4.00
C13	351.35	1.78	88.56	5	2	4	0.88	25	-0.83	1.00
C14	331.35	1.58	74.57	5	2	3	0.89	24	-0.72	0.39
C15	264.28	3.46	47.14	3	1	2	0.77	20	-0.43	20.00
...	...	...	...	...	...	...	...	...	...	...
C3218	382.43	4.10	97.49	7	1	3	0.42	26	-1.79	7.00
C3219	839.05	-7.90	408.98	22	16	18	0.06	58	0.69	20.00

ARTICLE IN PRESS

ARTICLE IN PRESS

This scenario tests whether ^{m}BI descriptors provide non-redundant information beyond standard physicochemical descriptors.

Because some targets also appear among predictor columns in the physicochemical descriptor set, strict leakage prevention was applied. Let $F^{(s)}$ denote the index set of predictors used in scenario s, and let c_t denote the index of the raw predictor column that coincides with target t (when such a column exists). For each target t under scenario s, the effective feature index set is

$$F_t^{(s)} = F^{(s)} \setminus \{c_t\}, \tag{7}$$

so that the model cannot trivially 'read' the target from the inputs and must instead infer it from correlated descriptors.

Descriptor redundancy and inter-correlation within ^{m}BI. The 16 modified bond-based indices form a structured family obtained by modulating related degree-based kernels; consequently, some inter-correlation among them is expected. In the present benchmark we do not assume that all 16 descriptors contribute independent information in a strict statistical sense. Instead, we evaluate their *net* predictive contribution through the three scenario comparisons (Modified bond-based indices only, Physicochemical only, and Combined). In particular, if the ^{m}BI set were largely redundant with itself and/or with the physicochemical panel, then concatenation would be unlikely to improve external-test performance under a broad model zoo. Conversely, systematic improvements under the Combined scenario provide direct evidence that the ^{m}BI family contributes complementary signal at the representation level, even if individual indices are partially correlated. Since our model zoo includes both regularized linear estimators and tree/boosting ensembles, the training stage is also not unduly sensitive to collinearity: regularization mitigates multicollinearity in linear models, while tree/boosting methods can exploit redundant predictors without requiring strict feature independence.

ARTICLE IN PRESS

4.4 Multi-task formulation and data splitting

The study is organised as a multi-task regression benchmark ,as explained in Figure 2, in which all targets share the same molecules and descriptor scenarios. For each target t and scenario s, a feature matrix and response vector are defined as

$$\mathbf{X}_t^{(s)} \in \mathbb{R}^{N \times d_s}, \qquad \mathbf{y}^{(t)} \in \mathbb{R}^N, \tag{8}$$

where d_s is the number of predictors under scenario s after leakage removal. To enforce strict comparability across all experiments, a single train–test split was defined at the molecule level and reused for every target and scenario. Let $I_{\text{train}} \subset \{1, \ldots, N\}$ denote the training indices and I_{test} the complementary test indices, with $|I_{\text{test}}| \approx 0.2N$. The resulting partitions are

$$\mathbf{X}_{t,\text{train}}^{(s)} = \mathbf{X}_t^{(s)}[I_{\text{train}}, :], \qquad \mathbf{X}_{t,\text{test}}^{(s)} = \mathbf{X}_t^{(s)}[I_{\text{test}}, :], \tag{9}$$

with analogous definitions for $\mathbf{y}_{\text{train}}^{(t)}$ and $\mathbf{y}_{\text{test}}^{(t)}$.

All predictor variables were standardised using statistics computed on the training set only. For feature j in scenario s, with training mean $\mu_j^{(s)}$ and standard deviation $\sigma_j^{(s)}$, the z-score normalised

ARTICLE IN PRESS

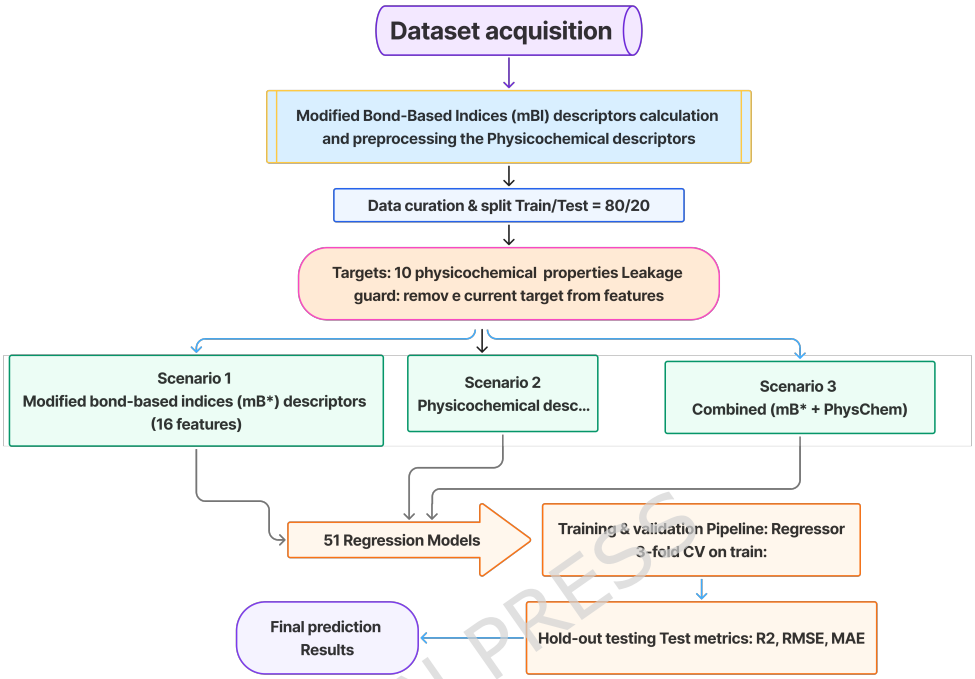

Figure 2: Workflow for constructing, benchmarking, and interpreting QSPR/QSAR models based on a curated set of 3219 antibacterial compounds.

ARTICLE IN PRESS

feature value is

$$\tilde{x}_{ij}^{(s)} = \frac{x_{ij}^{(s)} - \mu_j^{(s)}}{\sigma_j^{(s)}}, \tag{10}$$

with appropriate safeguards for features with zero variance. The same affine transformation is then applied to the test set, ensuring that no information from the test distribution is used during training.

4.5 Regression model families and hypothesis-testing strategy

To evaluate descriptor scenarios across a wide range of inductive biases, a broad "model zoo" of regression algorithms was benchmarked. The model set spans linear and regularised linear estimators, robust regression, kernel-based methods, nearest-neighbour regression, decision trees, bagging ensembles and gradient-boosting methods. Tree ensemble approaches play a central role due to their strong empirical performance on descriptor-based QSAR/QSPR. Random forest regression constructs an ensemble of decision trees trained on bootstrap samples with random feature subselection, with predictions obtained by averaging across trees [89]. Gradient boosting fits an additive sequence of weak learners (typically shallow trees) to iteratively reduce residual errors, providing a flexible framework for nonlinear regression [90]. In addition to scikit-learn implementations, modern gradient-boosting libraries were included as competitive baselines: XGBoost [91], LightGBM [92] and CatBoost [93]. Collectively, these methods provide a rigorous and diverse basis for testing whether performance improvements are consistent across model families rather than being artefacts

ARTICLE IN PRESS

of a particular algorithm.

4.6 Training protocol, cross-validation and performance metrics

For each target t, descriptor scenario s and regression model m, training was performed on the standardised training data and internal stability was assessed via three-fold cross-validation. After cross-validation, each model was retrained on the full training set and evaluated on the independent test set. For test samples $i \in I_{\text{test}}$, predictions $\hat{y}_i$ were compared to true values y_i using the coefficient of determination, root mean squared error and mean absolute error. The test-set coefficient of determination is

$$R^2_{t,s,m} = 1 - \frac{\sum_{i \in I_{\text{test}}} (y_i - \hat{y}_i)^2}{\sum_{i \in I_{\text{test}}} (y_i - \bar{y})^2}, \tag{11}$$

where $\bar{y}$ is the mean of $\{y_i : i \in I_{\text{test}}\}$. The corresponding error metrics are

$$\text{RMSE}_{t,s,m} = \sqrt{\frac{1}{|I_{\text{test}}|} \sum_{i \in I_{\text{test}}} (y_i - \hat{y}_i)^2}, \qquad \text{MAE}_{t,s,m} = \frac{1}{|I_{\text{test}}|} \sum_{i \in I_{\text{test}}} |y_i - \hat{y}_i|. \tag{12}$$

For each target and scenario, the best-performing model was selected as the regressor minimising test-set RMSE, and its corresponding R^2, RMSE and MAE values were retained for analysis and visualisation.

4.7 Multi-target metrics and descriptor scenario comparison

Beyond per-target evaluation, scenario performance was summarised across all $T = 10$ targets using macro-averaged and global metrics. Let $m^*_{t,s}$ denote the selected best model for target t under scenario s, and let $R^2_{t,s,m^*_{t,s}}$ denote its test-set coefficient of determination. The macro-averaged coefficient of determination is defined as

$$\text{Macro-}R^2(s) = \frac{1}{T} \sum_{t=1}^{T} R^2_{t,s,m^*_{t,s}}. \tag{13}$$

To compare targets on a common scale, a global z-RMSE was also computed by standardising each target dimension on the test set. Let μ_t and σ_t denote the mean and standard deviation of target t over the test set, and define

$$Z^{(s)}_{\text{true},i,t} = \frac{Y^{(s)}_{\text{true},i,t} - \mu_t}{\sigma_t}, \qquad Z^{(s)}_{\text{pred},i,t} = \frac{Y^{(s)}_{\text{pred},i,t} - \mu_t}{\sigma_t}. \tag{14}$$

The z-RMSE is then

$$\text{zRMSE}(s) = \sqrt{\frac{1}{|I_{\text{test}}|\, T} \sum_{i \in I_{\text{test}}} \sum_{t=1}^{T} \left(Z^{(s)}_{\text{true},i,t} - Z^{(s)}_{\text{pred},i,t}\right)^2}. \tag{15}$$

Together, these summaries provide a stringent multi-target comparison of the Modified bond-based indices descriptors (^{m}BI) scenario, the Physicochemical descriptors scenario and the Combined descriptors scenario, enabling assessment of whether observed gains reflect consistent improvements across the full set of predicted properties.

ARTICLE IN PRESS

4.8 Results and Discussion: Multi-task QSAR/QSPR Benchmark of ^{m}BI and Physicochemical Descriptors

Benchmark completion under a heterogeneous regression panel (hypothesis-driven design)

To test the stated hypothesis under minimal methodological bias, the evaluation was conducted as a large-scale benchmark over a deliberately heterogeneous panel of regression algorithms spanning multiple modelling families (linear and penalised regressors, robust estimators, kernel and support-vector methods, nearest-neighbour regression, decision trees, bagging/forest ensembles, and modern gradient-boosting frameworks). Under the fixed molecule-level train–test split and three-fold cross-validation protocol described in the Methods, a total of **1,270** target–scenario–model evaluations were successfully completed. This design is central to the hypothesis test: any observed descriptor advantage must be reproducible across diverse inductive biases rather than being contingent on a single algorithmic class.

ARTICLE IN PRESS

Overall multi-target performance across descriptor scenarios (primary hypothesis test)

The primary hypothesis concerns whether the mathematical descriptors add predictive value beyond standard physicochemical descriptors when assessed at the multi-target level. For each descriptor scenario, the *per-target best model* was selected using the predefined criterion (minimum external-test RMSE for that target within the scenario), and these best-per-target predictions were then aggregated into scenario-level multi-target summaries. Table 5 reports these results.

The **Physicochemical descriptors scenario** provides a strong baseline (Macro Test $R^2 = 0.852$, Global $R^2 = 0.970$, Global zRMSE $= 0.385$). The **Combined descriptors scenario** yields the best overall performance (Macro Test $R^2 = 0.861$, Global $R^2 = 0.972$, Global zRMSE $= 0.373$), corresponding to a reduction in global zRMSE of 0.012, i.e., $\approx 3.13\%$ relative to the Physicochemical descriptors scenario. In contrast, the **Modified bond-based indices descriptors (^{m}BI) scenario** is markedly weaker in aggregate (Macro Test $R^2 = 0.673$, Global $R^2 = 0.913$, zRMSE $= 0.572$). Overall, these scenario-level results support the hypothesis in a nuanced but direct sense: *the ^{m}BI descriptors are not sufficient as a stand-alone representation for this ten-endpoint panel, but they provide complementary information that improves multi-target generalisation when fused with physicochemical descriptors under a broad model-zoo evaluation.*

Table 5: Overall multi-target evaluation by descriptor scenario on the external test set. Macro RMSE/MAE average errors across targets with different units; zRMSE is unit-free and therefore the most directly comparable global error across endpoints.

Scenario	NumTargets	Macro Test R^2	Macro Test RMSE	Macro Test MAE	Global R^2	Global zRMSE
Modified bond-based indices descriptors (^{m}BI) scenario	10	0.673	7.771	4.988	0.913	0.572
Physicochemical descriptors scenario	10	0.852	4.255	2.542	0.970	0.385
Combined descriptors scenario	10	0.861	4.113	2.400	0.972	0.373

ARTICLE IN PRESS

Target-level best performance and best-in-class models (secondary hypothesis evidence)

At the endpoint level, most physicochemical properties are predicted with high external-test accuracy under the best configuration for that target. The strongest results include molecular weight ($R^2 = 0.973$), polar surface area ($R^2 = 0.968$), HBA ($R^2 = 0.949$), HBD ($R^2 = 0.924$), rotatable bonds ($R^2 = 0.951$), and QED ($R^2 = 0.915$); heavy-atom count approaches a ceiling ($R^2 = 0.993$). Natural-product likeness is moderately predictable (best $R^2 = 0.696$). Across all descriptor scenarios and model families, antibacterial potency (pMIC) remains the most difficult endpoint (best $R^2 = 0.348$, RMSE = 0.547). Table 6 reports the single best result per target (best scenario chosen by minimum external-test RMSE).

Table 6: Best result per target (best scenario chosen by minimum external test RMSE).

Target	Best scenario	Best model	Test R^2	Test RMSE	Test MAE
Rotatable Bonds	Combined descriptors scenario	ExtraTreesRegressor	0.951	1.291	0.797
AlogP	Physicochemical descriptors scenario	ExtraTreesRegressor	0.916	0.746	0.466
HBA	Physicochemical descriptors scenario	XGBoost	0.949	0.767	0.500
HBD	Physicochemical descriptors scenario	XGBoost	0.924	0.554	0.357
Heavy Atoms	Physicochemical descriptors scenario	HistGradientBoostingRegressor	0.993	0.847	0.341
Molecular Weight	Combined descriptors scenario	HistGradientBoostingRegressor	0.973	25.456	14.378
Np Likeness Score	Combined descriptors scenario	XGBoost	0.696	0.546	0.357
Polar Surface Area	Physicochemical descriptors scenario	ExtraTreesRegressor	0.968	10.118	6.569
QED Weighted	Physicochemical descriptors scenario	ExtraTreesRegressor	0.915	0.076	0.047
pMIC	Physicochemical descriptors scenario	HistGradientBoostingRegressor	0.348	0.547	0.400

ARTICLE IN PRESS

Model-zoo performance landscape (top models per scenario)

To connect the hypothesis test directly to the model-zoo philosophy, Tables 7–9 summarise the test-set R^2 values across the ten targets for a representative subset of top-performing models within each descriptor scenario; the complete set of results for all evaluated models across all three scenarios is provided in the Supplementary Material. These summaries show that strong performance is consistently achieved by non-linear ensemble and boosting families (e.g., ExtraTrees, histogram-based gradient boosting, and XGBoost), which is expected when descriptor–property relationships are interaction-driven. Within this regime, the Combined descriptors scenario exhibits the clearest signature of descriptor complementarity, particularly for structure-sensitive endpoints such as Rotatable Bonds and Np Likeness Score, aligning with the scenario-level improvement observed in Table 5.

5 Discussion

The benchmark was constructed to serve as a rigorous hypothesis test: whether the modified bond-based indices (^{m}BI) descriptors enhance QSAR/QSPR prediction when the conclusion must hold across heterogeneous regression families. The scenario-level results provide direct support for complementarity. Relative to the Physicochemical descriptors scenario, the Combined descriptors scenario improves Macro Test R^2 (0.861 vs. 0.852), improves Global R^2 (0.972 vs. 0.970), and reduces the unit-free Global zRMSE (0.373 vs. 0.385). Because these summaries are computed

ARTICLE IN PRESS

Table 7: Top models (test R^2) under the Modified bond-based indices descriptors (^{m}BI) scenario.

Model	RotB	AlogP	HBA	HBD	HeavyAtoms	MolWt	NpLike	PSA	QED	pMIC
GammaRegressor	–	–	–	–	0.94	0.86	–	–	0.46	–
RandomForestRegressor	0.90	0.50	0.71	0.50	0.99	0.94	0.56	0.68	0.77	0.15
ExtraTreesRegressor	0.91	0.49	0.70	0.49	0.99	0.94	0.56	0.67	0.77	0.08
XGBoost	0.89	0.44	0.70	0.50	0.99	0.94	0.52	0.67	0.77	0.08
HistGradientBoostingRegressor	0.89	0.46	0.70	0.43	0.99	0.94	0.52	0.65	0.77	0.14
LightGBM	0.89	0.49	0.71	0.42	0.99	0.94	0.54	0.66	0.76	0.10
BaggingRegressor	0.90	0.46	0.68	0.47	0.99	0.94	0.50	0.65	0.75	0.09
GradientBoostingRegressor	0.89	0.41	0.66	0.40	0.99	0.94	0.43	0.64	0.74	0.16
CatBoost	0.88	0.41	0.67	0.40	0.99	0.94	0.44	0.64	0.75	0.14
PoissonRegressor	0.65	–	0.52	0.28	0.95	0.87	–	0.51	0.46	–
LassoLarsIC	0.87	0.45	0.67	0.43	0.99	0.94	0.42	0.66	0.55	0.01
LassoLarsCV	0.87	0.45	0.67	0.43	0.99	0.94	0.42	0.66	0.54	0.01
TransformedTargetRegressor	0.87	0.44	0.67	0.41	0.99	0.94	0.42	0.66	0.55	0.03
LinearRegression	0.87	0.44	0.67	0.41	0.99	0.94	0.42	0.66	0.55	0.03
BayesianRidge	0.86	0.45	0.67	0.42	0.99	0.94	0.42	0.66	0.55	0.01
ARDRegression	0.86	0.45	0.67	0.42	0.99	0.94	0.42	0.66	0.54	0.01
RidgeCV	0.85	0.44	0.67	0.42	0.99	0.94	0.41	0.66	0.54	0.01
KNeighborsRegressor	0.87	0.38	0.60	0.40	0.98	0.93	0.41	0.59	0.71	-0.02
OrthogonalMatchingPursuitCV	0.82	0.44	0.66	0.41	0.99	0.94	0.38	0.66	0.54	0.01
Ridge	0.83	0.43	0.66	0.41	0.99	0.94	0.39	0.64	0.54	0.01

ARTICLE IN PRESS

Table 8: Top models (test R^2) under the Physicochemical descriptors scenario.

Model	RotB	AlogP	HBA	HBD	HeavyAtoms	MolWt	NpLike	PSA	QED	pMIC
ExtraTreesRegressor	0.92	0.92	0.95	0.92	0.99	0.96	0.62	0.97	0.91	0.33
XGBoost	0.85	0.91	0.95	0.92	0.99	0.97	0.60	0.96	0.91	0.34
HistGradientBoostingRegressor	0.89	0.88	0.94	0.90	0.99	0.97	0.58	0.96	0.90	0.35
RandomForestRegressor	0.86	0.89	0.94	0.91	0.99	0.97	0.59	0.96	0.90	0.33
KNeighborsRegressor	0.85	0.89	0.92	0.87	0.98	0.95	0.54	0.95	0.89	0.32
CatBoost	0.87	0.87	0.93	0.87	0.99	0.96	0.51	0.95	0.89	0.31
BaggingRegressor	0.84	0.88	0.93	0.91	0.99	0.96	0.52	0.96	0.89	0.24
GradientBoostingRegressor	0.86	0.86	0.92	0.88	0.99	0.96	0.44	0.95	0.87	0.25
GammaRegressor	–	–	–	–	0.93	0.88	–	–	0.55	–
MLPRegressor	0.88	0.90	0.92	0.89	0.99	0.70	0.53	0.95	0.86	0.26
PoissonRegressor	0.63	–	0.80	0.68	0.96	0.92	–	0.86	0.50	–
ARDRegression	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
LassoLarsIC	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
LassoLarsCV	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
TransformedTargetRegressor	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
LinearRegression	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
RidgeCV	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
BayesianRidge	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
LassoCV	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05
Ridge	0.76	0.81	0.89	0.82	0.99	0.95	0.27	0.95	0.61	0.05

ARTICLE IN PRESS

Table 9: Top models (test R^2) under the Combined descriptors scenario.

Model	RotB	AlogP	HBA	HBD	HeavyAtoms	MolWt	NpLike	PSA	QED	pMIC
XGBoost	0.94	0.91	0.95	0.92	0.99	0.97	0.70	0.97	0.91	0.33
ExtraTreesRegressor	0.95	0.91	0.94	0.91	0.99	0.97	0.65	0.97	0.91	0.31
HistGradientBoostingRegressor	0.93	0.89	0.94	0.90	0.99	0.97	0.66	0.96	0.91	0.33
RandomForestRegressor	0.93	0.89	0.94	0.91	0.99	0.97	0.62	0.96	0.90	0.34
BaggingRegressor	0.90	0.88	0.93	0.90	0.99	0.97	0.56	0.96	0.89	0.28
CatBoost	0.92	0.87	0.93	0.87	0.99	0.96	0.57	0.95	0.90	0.31
KNeighborsRegressor	0.89	0.87	0.90	0.82	0.99	0.96	0.55	0.94	0.89	0.31
GradientBoostingRegressor	0.92	0.86	0.92	0.87	0.99	0.96	0.49	0.95	0.88	0.25
PoissonRegressor	0.84	–	0.81	0.68	0.96	0.93	–	0.86	0.52	–
MLPRegressor	0.94	0.90	0.93	0.88	0.99	0.62	0.66	0.94	0.87	0.22
GammaRegressor	–	–	–	–	0.94	0.88	–	–	0.54	–
TransformedTargetRegressor	0.88	0.82	0.90	0.82	0.99	0.96	0.47	0.95	0.65	0.07
LinearRegression	0.88	0.82	0.90	0.82	0.99	0.96	0.47	0.95	0.65	0.07
LassoLarsIC	0.88	0.82	0.90	0.83	0.99	0.96	0.47	0.95	0.64	0.06
LassoLarsCV	0.88	0.82	0.90	0.83	0.99	0.96	0.47	0.95	0.64	0.06
ARDRegression	0.87	0.82	0.90	0.82	0.99	0.96	0.47	0.95	0.64	0.05
BayesianRidge	0.87	0.81	0.90	0.82	0.99	0.96	0.46	0.95	0.64	0.05
RidgeCV	0.87	0.81	0.90	0.82	0.99	0.96	0.46	0.95	0.64	0.05
Ridge	0.86	0.81	0.90	0.82	0.99	0.96	0.45	0.95	0.64	0.05
LassoCV	0.86	0.81	0.89	0.82	0.99	0.96	0.44	0.95	0.63	0.05

ARTICLE IN PRESS

after independent per-target model selection within a model zoo spanning multiple inductive biases, the advantage cannot be attributed to a particular regressor class; instead, it indicates that the information added by ^{m}BI is *usable* by multiple strong learners and is not redundant with the physicochemical panel.

At the endpoint level, the findings further clarify *where* the hypothesis is supported. For many classical physicochemical endpoints (e.g., heavy-atom count, molecular weight, polar surface area, HBA/HBD, QED), the Physicochemical descriptors scenario already yields very high performance, consistent with the strong redundancy and shared information content among global physicochemical descriptors under strict leakage control. In this regime, large absolute gains are structurally unlikely because several endpoints are close to a practical ceiling. The Combined descriptors scenario nevertheless achieves measurable improvements on structure-sensitive properties-notably Rotatable Bonds and Np Likeness Score-which are plausibly linked to graph connectivity and motif composition that compact topology-based indices are designed to capture. This pattern is consistent with the intended role of ^{m}BI as an *orthogonal* descriptor family that enriches the representation rather than replacing physicochemical descriptors.

Finally, antibacterial potency (pMIC) remains the limiting endpoint across all scenarios and model families (best $R^2 \approx 0.35$). In a heterogeneous ChEMBL-derived context, this persistent gap suggests that potency is only partially determined by global physicochemical summaries and compact topological indices, and that additional information sources or richer structure encodings would likely be required to close the remaining predictive gap. The overall conclusion is therefore twofold: the ^{m}BI descriptors alone are insufficient for the full ten-endpoint panel, but when combined with physicochemical descriptors they provide complementary signal that yields consistent, multi-target improvements under a rigorous model-zoo hypothesis-testing strategy.

ARTICLE IN PRESS

6 Conclusions and Future Directions

In this work, we introduced and systematically analysed a broad family of *modified bond-based indices* on molecular graphs, obtained by coupling classical degree–based kernels with a normalising bond factor depending on the degrees of the incident vertices. The resulting descriptors ${}^{m}\mathrm{BM}_1$, ${}^{m}\mathrm{BM}_2$, ${}^{m}\mathrm{BF}$, ${}^{m}\mathrm{BY}$, ${}^{m}\mathrm{BPC}$, ${}^{m}\mathrm{BSC}$, ${}^{m}\mathrm{BN}$, ${}^{m}\mathrm{BABC}$, ${}^{m}\mathrm{BCAB}$, ${}^{m}\mathrm{BGO}_1$, ${}^{m}\mathrm{BGO}_2$, ${}^{m}\mathrm{BGA}$, ${}^{m}\mathrm{BH}$, ${}^{m}\mathrm{BMP}$, ${}^{m}\mathrm{BS}$, ${}^{m}\mathrm{BD}$ provide a unified framework that interpolates between vertex–centric and bond–centric information, while preserving the strictly local, chemically interpretable nature of degree–based indices.

On the structural side, we established an edge–partition representation for all modified bond-based indices in terms of degree pairs and bond degrees. This representation allowed us to derive closed forms for a wide range of benchmark graph families, including paths, cycles, complete graphs, complete bipartite graphs, stars, friendship graphs, wheel graphs, book graphs, Dutch windmill graphs and hypercubes. The resulting tables highlight the different growth regimes of the indices across sparse, tree-like and highly connected topologies, and identify natural extremal candidates within these classes. In addition, we obtained degree–extreme bounds for selected modified indices, expressed in terms of the order, size and minimum/maximum degree, which can be used as quick diagnostic tools or consistency checks in applications.

On the applied side, we complemented the analytical development with a stringent, hypothesis-driven QSAR/QSPR benchmark to test whether the proposed modified bond-based indices provide predictive value beyond standard physicochemical descriptors. Using a curated ChEMBL antibacterial dataset of $N = 3{,}219$ unique molecules and a ten-endpoint multi-task panel (nine physicochemical targets together with antibacterial potency encoded as pMIC), we compared three descriptor scenarios: the Modified bond-based indices descriptors (${}^{m}\mathrm{BM}$) scenario, the Physicochemical descriptors scenario, and the Combined descriptors scenario. This evaluation was conducted under a large and heterogeneous regression model zoo spanning linear, robust, kernel-based, nearest-neighbour, and ensemble/boosting regressors, while enforcing a fixed molecule-level train–test split and strict leakage control. At the multi-target level, the Combined descriptors scenario achieved the best overall generalisation, improving upon the Physicochemical descriptors scenario (Macro Test $R^2 = 0.861$ vs. 0.852; Global zRMSE $= 0.373$ vs. 0.385), whereas the Modified bond-based indices descriptors (${}^{m}\mathrm{BM}$) scenario alone remained substantially weaker in aggregate. These results support the central practical conclusion that the modified indices encode complementary structural signal that is most effective when fused with physicochemical descriptors and exploited by flexible non-linear learners.

Looking forward, several extensions are natural. First, the remaining performance gap for potency (pMIC) motivates incorporating richer structure encodings (e.g., fingerprints or learned molecular embeddings) and modelling assay heterogeneity more explicitly (e.g., organism- or assay-aware stratification), while retaining ${}^{m}\mathrm{BI}$ as an interpretable structural component. Second, multi-output learning strategies that exploit cross-target correlations directly (rather than per-target model selection) may further enhance global profile prediction, particularly when combined with uncertainty-aware evaluation. Finally, the edge-partition framework developed here suggests principled avenues for designing new bond-normalised indices (including higher-order neighbourhood variants) and for embedding such indices as structured inputs to graph-based machine-learning models, enabling a tighter integration between analytical graph descriptors and modern predictive chemoinformatics.

ARTICLE IN PRESS

Conflict of interest

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

Funding

No funding is available for this study.

Author contributions

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

Data Availability Statement

The data analyzed in this study are publicly available from the ChEMBL database (`https://www.ebi.ac.uk/chembl/`). The processed dataset is included in the Supplementary Information. Scripts used for descriptor computation and model benchmarking are available from the corresponding author upon reasonable request.

ARTICLE IN PRESS

ARTICLE IN PRESS

References

[1] M. V. Diudea, I. Gutman, & L. Jantschi. (2001). *Molecular Topology*. Huntington, NY.

[2] S. Wagner, & H. Wang. (2018). *Introduction to Chemical Graph Theory*. CRC Press.

[3] H. Wiener. (1947). Structural determination of the paraffin boiling points. *Journal of the American Chemical Society*, *69*, 17–20.

[4] H. Hosoya. (1971). Topological index: A newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons. *Bulletin of the Chemical Society of Japan*, *44*, 2332–2339.

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

[6] S. Nikolic, G. Kovacevic, A. Milicevic, & N. Trinajstic. (2003). The Zagreb indices 30 years after. *Croatica Chemica Acta*, *76*, 113–124.

[7] I. Gutman, & K. C. Das. (2004). The first Zagreb index 30 years after. *MATCH Communications in Mathematical and Computer Chemistry*, *50*, 83–92.

[8] K. Xu, M. Liu, K. C. Das, I. Gutman, & B. Furtula. (2017). A survey on Zagreb indices. *MATCH Communications in Mathematical and Computer Chemistry*, *78*, 17–56.

[9] M. Randić. (1975). On characterization of molecular branching. *Journal of the American Chemical Society*, *97*, 6609–6615.

[10] M. Randić. (2001). The connectivity index 25 years after. *Journal of Molecular Graphics and Modelling*, *20*, 19–35.

[11] X. Li, & Y. Shi. (2008). A survey on the Randić index. *MATCH Communications in Mathematical and Computer Chemistry*, *59*, 127–156.

[12] R. Todeschini, & V. Consonni. (2008). *Handbook of Molecular Descriptors*. Wiley.

[13] D. J. Klein. (2002). Topological indices and related descriptors in QSAR and QSPR (Review of Devillers & Balaban, 1999). *Journal of Chemical Information and Computer Sciences*, *42*(6), 1507–1507.

[14] T. Alraqad, A. Alameri, M. Alsharafi, & et. al. (2024). Analysis of octane isomer properties via topological descriptors of line graphs. Scientific Reports, 14(1), 27159. https://doi.org/10.1038/s41598-024-75454-6

[15] S. C. Basak, B. D. Gute, & G. D. Grunwald. (1996). Estimation of the normal boiling points of haloalkanes using molecular similarity. *Croatica Chemica Acta*, *69*(3), 1159–1173.

[16] P. V. Khadikar. (2000). On a novel structural descriptor PI. *National Academy Science Letters*, *23*, 113–118.

ARTICLE IN PRESS

[17] E. Estrada, L. Torres, L. Rodríguez, & I. Gutman. (1998). An atom–bond connectivity index: Modelling the enthalpy of formation of alkanes. *Indian Journal of Chemistry, Section A*, *37A*, 849–855.

[18] B. Zhou. (2004). Zagreb indices. *MATCH Communications in Mathematical and Computer Chemistry*, *52*, 113–119.

[19] B. Zhou, & I. Gutman. (2005). Further properties of Zagreb indices. *MATCH Communications in Mathematical and Computer Chemistry*, *54*, 233–239.

[20] S. Vaidya, & J. Chang. (2024). Sharp bounds for generalized Zagreb indices of graphs. *arXiv preprint arXiv:2409.06081*.

[21] J. Li, Y. Shi, & I. Gutman. (2020). Nordhaus–Gaddum-type results for Zagreb indices. *Discrete Applied Mathematics*, *284*, 396–402.

[22] I. Gutman, & B. Furtula (Eds.). (2008). *Recent Results in the Theory of Randić Index*. Math. Chem. Monogr. 6, University of Kragujevac.

[23] X. Li, & I. Gutman. (2006). *Mathematical Aspects of Randić-Type Molecular Structure Descriptors*. University of Kragujevac.

[24] K. C. Das. (2010). On geometric-arithmetic index of graphs. *MATCH Communications in Mathematical and Computer Chemistry*, *64*, 619–630.

[25] Y. Yuan, B. Zhou, & N. Trinajstić. (2010). On geometric-arithmetic index. *Journal of Mathematical Chemistry*, *47*, 833–841.

[26] Z. Du, B. Zhou, & N. Trinajstić. (2011). On geometric-arithmetic indices of (molecular) trees, unicyclic graphs and bicyclic graphs. *MATCH Communications in Mathematical and Computer Chemistry*, *66*, 681–697.

[27] L. Zhong. (2012). The harmonic index for graphs. *Applied Mathematics Letters*, *25*, 561–566.

[28] M. Cheng, & L. Wang. (2016). A lower bound for the harmonic index of a graph with minimum degree at least three. *Filomat*, *30*(8), 2249–2260.

[29] A. Ali, L. Zhong, & I. Gutman. (2019). Harmonic index and its generalization: Extremal results and bounds. *MATCH Communications in Mathematical and Computer Chemistry*, *81*, 249–311.

[30] E. Estrada. (2008). Atom-bond connectivity and the energetics of branched alkanes. *Chemical Physics Letters*, *463*, 422–425.

[31] B. Furtula, A. Graovac, & D. Vukičević. (2009). Atom-bond connectivity index of trees. *Discrete Applied Mathematics*, *157*, 2828–2835.

[32] B. Zhou, & N. Trinajstić. (2009). On a novel connectivity index. *Journal of Mathematical Chemistry*, *46*, 1252–1270.

ARTICLE IN PRESS

ARTICLE IN PRESS

[33] A. Ali, K. C. Das, D. Dimitrov, & B. Furtula. (2021). Atom-bond connectivity index of graphs: A review over extremal results and bounds. *Discrete Mathematics Letters*, *5*, 68–93.

[34] B. Zhou, & N. Trinajstić. (2009). On general sum-connectivity index. *Journal of Mathematical Chemistry*, *46*, 283–293.

[35] Y. Tang, D. B. West, & B. Zhou. (2016). Extremal problems for degree-based topological indices. *Discrete Applied Mathematics*, *203*, 134–143.

[36] I. Gutman. (2021). Geometric approach to degree-based topological indices. *MATCH Communications in Mathematical and Computer Chemistry*, *86*, 11–16.

[37] A. Gagarin, K. Chellali, & O. D. Makinde. (2022). Generalized Zagreb indices and their applications. *Discrete Mathematics Letters*, *8*, 30–38.

[38] B. Furtula, & I. Gutman. (2015). A forgotten topological index. *Journal of Mathematical Chemistry*, *53*, 1184–1190.

[39] Z. Che, & Z. Chen. (2016). Lower and upper bounds of the forgotten topological index. *MATCH Communications in Mathematical and Computer Chemistry*, *76*, 635–648.

[40] J. B. Liu, M. M. Matejić, E. I. Milovanović, & I. Ž. Milovanović. (2020). Some new inequalities for the forgotten topological index and coindex of graphs. *MATCH Communications in Mathematical and Computer Chemistry*, *84*, 719–738.

[41] B. Chaluvaraju, H. S. Boregowda, & I. N. Cangul. (2021). Some inequalities for the first general Zagreb index of graphs and line graphs. *Proceedings of the National Academy of Sciences, India, Section A: Physical Sciences*, *91*, 79–88.

[42] H. Abdo, D. Dimitrov, & I. Gutman. (2017). On extremal trees with respect to the F-index. *Kuwait Journal of Science*, *44*(3).

[43] C. Natarajan, S. K. Ayyaswamy, & D. Sarala et al. (2021). On F-index of certain generalized thorny graphs. *Proceedings of the National Academy of Sciences, India, Section A: Physical Sciences*, *91*, 269–272.

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

[45] W. Gao, M. R. Farahani, & L. Shi. (2016). Forgotten topological index of some drug structures. *Acta Medica Mediterranea*, *32*(1), 579–585.

[46] S. Mondal, N. De, & A. Pal. (2019). On some general neighborhood degree based topological indices. *International Journal of Applied Mathematics*, *32*(6), 1037–1056.

[47] S. Mondal, N. De, & A. Pal. (2021). On neighborhood Zagreb index of product graphs. *Journal of Molecular Structure*, *1223*, 129210.

[48] H. Ahmed, A. Saleh, R. Ismail, R. Salestina M, & A. Alameri. (2023). Computational analysis for eccentric neighborhood Zagreb indices and their significance. *Heliyon*, *9*, e17998.

ARTICLE IN PRESS

ARTICLE IN PRESS

[49] K. C. Das, & S. Mondal. (2023). On neighborhood inverse sum indeg index of molecular graphs with chemical significance. *Information Sciences*, *623*, 112–131.

[50] N. Benakli, E. Halleck, & S. R. Kingan. (2019). A characterization of 2-neighborhood degree list of diameter 2 graphs. *arXiv preprint arXiv:1909.08722*.

[51] D. Balasubramaniyan, & N. Chidambaram. (2023). On some neighbourhood degree-based topological indices with QSPR analysis of asthma drugs. *The European Physical Journal Plus*, *138*, 823.

[52] S. Hayat, M. Adil Khan, & et. al. (2023). Extremal hyper-Zagreb index of trees of given segments with applications to regression modeling in QSPR studies, textitAlexandria Engineering Journal, textit80, 259–268.

[53] S. Vaidya, & G. Surendran. (2016). Bounds for topological indices of molecular graphs. *Congressus Numerantium*, *227*, 51–64.

[54] X. Zhu, Q. Huang, & H. Deng. (2020). New generalizations of the Zagreb indices and their extremal properties. *Journal of Mathematical Chemistry*, *58*, 1095–1114.

[55] M. Tavakoli, F. Rahbarnia, & A. R. Ashrafi. (2014). Some new results on irregularity of graphs. *Journal of Applied Mathematics & Informatics*, *32*, 675–685.

[56] H. Wang, & S. Yuan. (2016). On the sum of squares of degrees and products of adjacent degrees. *Discrete Mathematics*, *339*, 1212–1220.

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

[58] S. Hayat, S.J.F. Alenazi, J.-B. Liu. (2024). 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*, 055027.

[59] S. Hayat, &-B. Liu. (2025). 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), 2550047.

[60] A. Ali, S. Elumalai, & T. Mansour. (2020). On the symmetric division deg index of molecular graphs. *MATCH Communications in Mathematical and Computer Chemistry*, *83*, 205–220.

[61] B. Furtula, K. C. Das, & I. Gutman. (2018). Comparative analysis of symmetric division deg index as potentially useful molecular descriptor. *International Journal of Quantum Chemistry*, *118*, e25659.

[62] K. C. Das, M. Matejić, E. Milovanović, & I. Milovanović. (2019). Bounds for symmetric division deg index of graphs. *Filomat*, *33*, 683–698.

[63] A. Vasilyev. (2014). Upper and lower bounds of symmetric division deg index. *Iranian Journal of Mathematical Chemistry*, *5*(2), 91–98.

ARTICLE IN PRESS

[64] A. M. Albalahi, & A. Ali. (2022). On the inverse symmetric division deg index of unicyclic graphs. *Computation*, *10*, 181.

[65] Z. Raza, L. Saha, & K. C. Das. (2023). On inverse symmetric division deg index of graphs. *RAIRO - Operations Research*, *57*, 3223–3236.

[66] L. Buyantogtokh, B. Horoldagva, & K. C. Das. (2022). On general reduced second Zagreb index of graphs. *Mathematics*, *10*, 3553.

[67] M. Ghorbani, S. Zangi, & N. Amraei. (2021). New results on symmetric division deg index. *Journal of Applied Mathematics and Computing*, *65*, 161–176.

[68] D. Vukičević. (2010). Bond additive modeling 1: Adriatic indices. *Croatica Chemica Acta*, *83*, 243–260.

[69] D. Vukičević. (2010). Bond additive modeling 2. Mathematical properties of max–min rodeg index. *Croatica Chemica Acta*, *83*, 261–273.

[70] D. Vukičević, & B. Furtula. (2009). Topological index based on the ratios of geometrical and arithmetical means of end-vertex degrees of edges. *Journal of Mathematical Chemistry*, *46*, 1369–1376.

[71] J. Du, & X. Sun. (2024). On bond incident degree index of chemical trees with a fixed order and a fixed number of leaves. *Applied Mathematics and Computation*, *464*, 128390.

[72] A. Ali, B. Furtula, I. Redzepović, & I. Gutman. (2022). Atom-bond sum-connectivity index. *Journal of Mathematical Chemistry*, *60*, 2081–2093.

[73] A. Ali, I. Gutman, & I. Redžepović. (2023). Atom-bond sum-connectivity index of unicyclic graphs and some applications. *Electronic Journal of Mathematics*, *5*, 1–7.

[74] E. Estrada. (2000). Characterization of 3D molecular structure. *Chemical Physics Letters*, *319*, 713–718.

[75] B. Zhou. (2008). On Estrada index. *MATCH Communications in Mathematical and Computer Chemistry*, *60*(2), 485–492.

[76] I. Gutman, B. Furtula, V. Marković, & B. Glišić. (2007). Alkanes with greatest Estrada index. *Zeitschrift für Naturforschung A*, *62*(9), 495–498.

[77] N. Soleimani, M. J. Nikmehr, & H. A. Tavallaee. (2015). Computation of the different topological indices of nanostructures. *Journal of the National Science Foundation of Sri Lanka*, *43*(2).

[78] H. Yousefi-Azari, & A. R. Ashrafi. (2008). Padmakar–Ivan index of q-multi-walled carbon nanotubes and nanotori. *Journal of Computational and Theoretical Nanoscience*.

[79] M. Alsharafi, M. Shubatah, & A. Alameri. (2020). The hyper-Zagreb index of some complement graphs. *Advances in Mathematics: Scientific Journal*, *9*, 3631–3642.

ARTICLE IN PRESS

ARTICLE IN PRESS

[80] M. Alsharafi, A. Alameri, Y. Zeren, M. Shubatah, & A. Alwardi. (2024). The Y-index of some complement graph structures and their applications to nanotubes and nanotorus. *Journal of Mathematics*, *2024*, 4269325.

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

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

[83] L. R. M. Gnanaraj, D. Ganesan, & M. K. Siddiqui. (2023). Topological indices and QSPR analysis of NSAID drugs. *Polycyclic Aromatic Compounds*, 1–17.

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

[85] R. Manonmani, & P. Selvarani. (2025). Multilinear regression analysis of antiviral medications with topological indices. *Journal of Molecular Graphics and Modelling*, *140*, 109104.

[86] S. Mondal, N. De, & A. Pal. (2022). Topological indices of some chemical structures applied for the treatment of COVID-19 patients. *Polycyclic Aromatic Compounds*, *42*(4), 1220–1234.

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

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

[89] L. Breiman. (2001). Random forests. *Machine Learning*, *45*(1), 5–32. https://doi.org/10.1023/A:1010933404324

[90] J. H. Friedman. (2001). Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*, *29*(5), 1189–1232. https://doi.org/10.1214/aos/1013203451

[91] T. Chen, & C. Guestrin. (2016). XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (pp. 785–794). https://doi.org/10.1145/2939672.2939785

[92] G. Ke, Q. Meng, T. Finley, T. Wang, W. Chen, W. Ma, Q. Ye, & T.-Y. Liu. (2017). LightGBM: A highly efficient gradient boosting decision tree. In *Advances in Neural Information Processing Systems*, *30*.

[93] L. Prokhorenkova, G. Gusev, A. Vorobev, A. V. Dorogush, & A. Gulin. (2018). CatBoost: Unbiased boosting with categorical features. In *Advances in Neural Information Processing Systems*, *31*.