MATCH
Communications in Mathematical and in Computer Chemistry

MATCH Commun. Math. Comput. Chem. **90** (2023) 743–763
ISSN: 0340–6253
doi: 10.46793/match.90-3.743T

Relationship between Ordinary, Laplacian, Randić, Incidence, and Sombor Energies of Trees

Hafsah Tabassum[a,b], Pawaton Kaemawichanurat[a,b], Adeela[c], Nathakhun Wiroonsri[a,b,*]

[a] *Department of Mathematics, Faculty of Science, King Mongkut's University of Technology Thonburi, Bangkok, Thailand*

[b] *Mathematics and Statistics with Applications (MaSA)*

[c] *University of Management and Technology Lahore, Pakistan*

hafsahtabassum@yahoo.com, pawaton.kae@kmutt.ac.th, adeela.abdulhameed@gmail.com, nathakhun.wir@kmutt.ac.th

(Received May 7, 2023)

Abstract

The sum of the absolute values of the eigenvalues of the graph's adjacency matrix is known as its ordinary energy. Based on the eigenvalues of a range of other graph matrices, several other equivalent energies are being considered. In this work, we considered ordinary energy, Laplacian, Randić, incidence, and Sombor energy to analyze their relationship using polynomial regression. The performance of each model is exceptional with cross-validation RMSE mostly below 1.

1 Introduction

Quantitative structure–property relationship (QSPR) and Quantitative structure-activity relationship(QSAR) modeling, which are used regularly

*Corresponding author.

in many areas of chemistry, rely heavily on molecular descriptors. To this end, topological indices are frequently used due to their computational simplicity and ease of use in identifying structure-property connections. Topological descriptors exist in the hundreds. They can be conveniently sorted in accordance with the origin of the parameters employed in their definitions. Hence, there are topological molecular descriptors that are based on degree, distance, and eigenvalue, although there are a few that do not neatly fit into any of these categories. In this paper, our focus is on some topological descriptors based on eigenvalues. Since, the understanding of the physical significance of eigenvalues in HMO theory, interest in molecular topological descriptors based on eigenvalues grew.

This emergence begins in the 1970s of the past century. Since then, the eigenvalue-based descriptors become a very well-researched topic that they are considered to be a separate branch of graph theory i.e. graph spectral theory(or algebraic graph theory). For each symmetric graph matrix, graph energies are matrix energies of different graph forms. They are called spectral indices in the field of quantitative chemistry. Spectral indexes might be single eigenvalues or functions of a matrix's eigenvalues known as the spectrum of the matrix.

A huge number of articles and monographs were written on the graph energies, for reference see [10,11]. An excellent collection of different results related to graph energy, its background, and the significance of the topic and methodologies that can serve as a guide for researchers interested in getting more knowledge can be found in the monograph [12]. Gutman and Frutella [14] pointed out the main applications along with the basic facts and principles to study graph energies. In [13], the bounds for the graph energy were described by Gutman, however, McClelland was the one to estimate bounds for the graph energy for the first time in 1971 [19]. Later on, Gutman with Koolen and Moulton introduced a modification of McClelland's formula for total π-electron energy.

The comprehensive research on the topic began 25 years after the introduction of graph energy. More than a hundred variants of graph energies have been defined using different matrices beyond the adjacency matrix.

Several graph invariants depending on the degrees of vertices can be found in the mathematical and chemical literature [9, 16]. One can differentiate between energies defined over different adjacency matrices, the Incidence energy, and Sombor energy can be defined on the incidence and Sombor matrix respectively. Likewise, Laplacian and Randić energy is defined over the Laplacian and Randić matrices respectively. Each of these energies can be quantified in greater depth by employing a specific functional formula depending upon the collection of eigenvalues. Typically, matrix energy is the sum of the absolute eigenvalues of the simple adjacency matrix but other variants can be found. One can, for example, use the sum of minimum or maximum eigenvalue, positive eigenvalues, and the difference between the greatest and smallest eigenvalue i.e. the diameter of a spectrum, etc. We would like to skip the details here, anyone interested can see [3, 17] for details.

More than a thousand publications and on average two papers per week (according to research statistics) show graph energies as an active research topic nowadays. The main reason behind this unexpected growth is unexpected applications of the graph energies in different fields of engineering and science [5, 8] that include air transportation [15], face recognition [1], protein sequence comparison [6], high satellite resolution [2], spacecraft construction, crystallography [25], complex network [24], etc. Some other applications have also been noticed in medicine. In 2009 Stevanović, D. and Stanković, I. studied the relationship between simple and Laplacian energy variants of graphs [22]. Mikołaj Morzy et al. found the relationship of different energies and their relationship with different centrality measures for egocentric networks and [20]. Shao, Yanling, et al. studied some upper and lower bounds of degree-based energies of trees [21].

This work aims to investigate the relationship of some graph energies of arbitrary tree graphs. We undertake a comparative analysis of some eigenvalue-based topological indices i.e. graph energies by using machine learning algorithms.

The paper is organized in the following manner: Next section contains some preliminaries such as the types of energies we are considering in this

work. The methodology and computational details are presented in Section 3. The relationship of all five energies with each other was examined. Finally, based on our data set the graph energies are compared. Section 4 includes some results and discussion. Section 5 contains some concrete examples of the models produced by the statistical techniques employed and the last section gives the conclusion of our work.

2 Preliminaries

Let $V(G)$ be a vertex set of an un-directed graph G and $E(G)$ be an edge set. Let n and m be the number of vertices and the number of edges respectively. If the vertices u and $v \in V(G)$ are adjacent then uv denotes the edge between these vertices. d_v represents the degree of the vertex v and can be defined as the number of edges incident to the vertex v. Graph energy is defined as the sum of the absolute value of eigenvalues of a graph G given by $E(G) = \sum_{j=1}^{n} |\lambda_j|$. Let $\lambda_1, \lambda_2, \ldots, \lambda_n$ be its eigenvalues [8], for each $1 \leq j \leq n$, λ_j be the roots of the characteristic polynomial $\phi(G; x) = det(xI - A(G))$ where $A(G)$ represents adjacency matrix of graph G. For the graph G having vertex set $v_1, v_2, \ldots, v_n$, its adjacency matrix $A(G) = (a_{ij})_{n \times n}$ is an n order symmetric matrix, whose elements are defined in [7] as:

$$a_{ij} = \begin{cases} 1 & \text{if } v_i v_j \in E(G) \\ \\ 0 & \text{if } v_i v_j \notin E(G) \end{cases}$$

$$TI(G) = \sum_{uv \in E(G)} \Phi(d_u, d_v)$$

where Φ is suitable function with the condition $\Phi(x, y) = \Phi(y, x)$. The graph invariants stated above are known as topological indices. *Laplacian*

matrix $L_{i,j}$ can be defined as:

$$L_{ij} = \begin{cases} deg(v_i) & \text{if } i = j \\ -1 & \text{if } i \neq j \text{ and } v_i v_j \in E(G) \\ 0 & \text{otherwise} \end{cases}$$

Let λ_i for $1 \leq i \leq n$ be the eigenvalues of Laplacian matrix. The *Laplacian energy* can be defined as:

$$LE(G) = \sum_{i=1}^{n} |\lambda_i - \frac{2m}{n}|$$

where m is cardinality of edge set and n be cardinality of vertex set of G. The Randić matrix [4] given by $R(G) = (r_{ij})_{n \times n}$ can be defined as:

$$r_{ij} = \begin{cases} \frac{1}{\sqrt{d_i d_j}} & \text{if } v_i v_j \in E(G) \\ 0 & \text{if } v_i v_j \notin E(G) \end{cases}$$

The sum of absolute values of the eigenvalues of this Randić matrix is known as the Randić energy $RE(G)$ of the graph.

The incidence matrix $I(G)$ of an undirected graph G has a column for each edge and a row for each vertex of the graph.

$$I_{ij} = \begin{cases} 1 & \text{if vertex } v_i \text{ is incident to edge } e_j \\ 0 & otherwise \end{cases}$$

Incidence energy is the sum of the singular values of the incidence matrix $I(G)$ that are, in turn, equal to the square root of eigenvalues of $I(G)I(G)^t$ where $I(G)I(G)^t$ is a square matrix of order n.

The Sombor Index $SO(G)$, developed by Gutman [14], is a novel graph invariant. This index is noticeable because of its connection to well-known degree-based topological indices known as forgotten indices. The Sombor

matrix, denoted by $A_{SO}(G) = (so_{ij})_{n\times n}$, of the graph G is a symmetric matrix having order n with the following elements:

$$so_{ij} = \begin{cases} \sqrt{d_i^2 + d_j^2} & \text{if } v_iv_j \in E(G) \\ \\ 0 & \text{if } v_iv_j \notin E(G) \end{cases}$$

The sum of the absolute value of eigenvalues of the Sombor matrix so_{ij} is known as the Sombor energy $SE(G)$ of the graph.

A connected graph with no cyclic subgraph is known as a *tree*. The term *leaf* or *pendent vertex* refers to a vertex having degree one. The incident edge of a leaf is referred to as the *pendent edge*, while its neighboring vertex is referred to as the *support vertex*. A tree T with one vertex designated as the root r is called a *rooted tree*. *Polynomial regression* is a type of regression model in which an n-th order polynomial represents the relationship between the independent and dependent variables. Consequently, it can be used to characterize a nonlinear relationship between X and y.

Statistically, polynomial regression applies a nonlinear model to the data, but the regression is basically linear [18]. Generally, it can be written as:

$$y = \beta_0 + \vec{\beta_1}^{\,t}\vec{X} + \vec{\beta_2}^{\,t}\vec{X^2} + \cdots + \vec{\beta_d}^{\,t}\vec{X^d} + \epsilon$$

where β_0 is a constant, $\vec{X} = (X_1, X_2, \ldots, X_p)$ and $\vec{\beta_k} = (\beta_{k1}, \ldots, \beta_{kp})$ are p-dimensional vectors, $k = 1, 2, \ldots, d$, with d be the maximum degree, and $\vec{\beta_k}$ represents the coefficients of the features considered in the model and ϵ is the error term. When we do linear regression to predict a statistically significant relationship between the predictor and response variable. The regression coefficients are tested for significance using the t-test. The following hypotheses are used for this test:

- H_0: $\beta_{kj} = 0$ (X_j^k does not affect the response (y))
- H_A: $\beta_{kj} \neq 0$ (X_j^k affects the response (y))

known as null and alternative hypotheses respectively. The test statistic

can be calculated as follows:

$$t = \hat{\beta}_{kj}/SE_{kj} \tag{1}$$

where:

- $\hat{\beta}_{kj}$: coefficient estimate of β_{kj} for X_j^k
- SE_{kj}: standard error of the coefficient estimate of β_{kj} for X_j^k

A statistical measure used to evaluate the strength of the evidence against the null hypothesis while testing the hypothesis is known as the p-value. If the p-value associated with t is below some threshold (e.g. $\beta = 0.05$), then we reject the null hypothesis and draw a conclusion that there is a statistically significant relationship between the predictor and response variables. The degree and direction of the relationship between the independent and dependent variables in a regression model can be assessed by β.

In this work, we considered up to third-degree polynomial. To assess how good a trained model is at making predictions, one must compare the actual values and the model's predictions. Therefore, it is important to look at some statistical measures to see how well the model can predict. The best predictive model will of course record the least error. So, in our case, we are looking at *RMSE* i.e. Root Mean Square Error that is given as:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{n}(\hat{y}_t - y_t)^2}.$$

When assessing a model's efficacy in machine learning and data science, k-fold cross-validation ($k = 10$ in our case) is frequently employed. By dividing the data into k equal-sized chunks, we may train the model on $k-1$ partitions and then validate it on the last partition. This procedure is done a total of k times, with each partition being validated only once. k-fold cross-validation estimates the model's generalization performance by averaging the performance measures across k-folds. It is a useful technique for estimating the model's efficacy on unseen data and preventing over- and under-fitting. In this paper, we use k-folds to estimate RMSE for the number of vertices above 11. Although this approach is reliable, there

are instances when it should not be employed, such as when you have a very limited dataset to analyze. So, for $6 \leq n \leq 10$ we use Leave-one-out cross-validation to estimate RMSE. Leave-one-out cross-validation is used when we have a very small dataset and it is a special case of k-folds with $k = n$. It is a way to figure out how well machine learning methods work when they are used to make predictions based on data that wasn't used to train the model. It takes a lot of computing power to do this process but it gives a reliable and unbiased estimate of model performance.

3 Methodology

This section includes the outcomes of the experimental analysis of relationships between energies. The methodology for analysis goes as given in Fig. 2 to investigate how the different energies of tree graphs are related. The trees considered in this work are all trees on n-vertices where $n = 6, 7, \ldots, 20$, that can be generated using the Python module Networkx. The number of trees corresponding to each n are given in the Table 1 below:

Table 1. Number of trees on n-vertices

n	**Number of Trees**
6	6
7	11
8	23
9	47
10	106
11	235
12	551
13	1301
14	3359
15	7741
16	19320
17	48629
18	123867
19	317955
20	823065

Our data set consists of five types of energy measures, and we created five instances of a model by changing one parameter. From the scatter plots in Fig. 1, a linear model may not be the best algorithm for comparing the energies under consideration. First, we scaled the independent variables to interpret them on the same scale. Then we applied polynomial regression by fitting a linear, quadratic, and cubic polynomial with k-fold cross-validation for $n = 11, 12, \ldots, 20$ and leave-one-out cross-validation for $n = 6, 7, \ldots, 10$ to estimate the root mean squared error. We generated 81 models for each model and every n with varying degree combinations. Out of which, the model with the lowest RMSE was selected to move forward. The selected models appeared overly complicated. The 'one-sigma' technique [23] was used to select the best model for each energy measure, with the smallest RMSE within one standard deviation from the minimum RMSE.

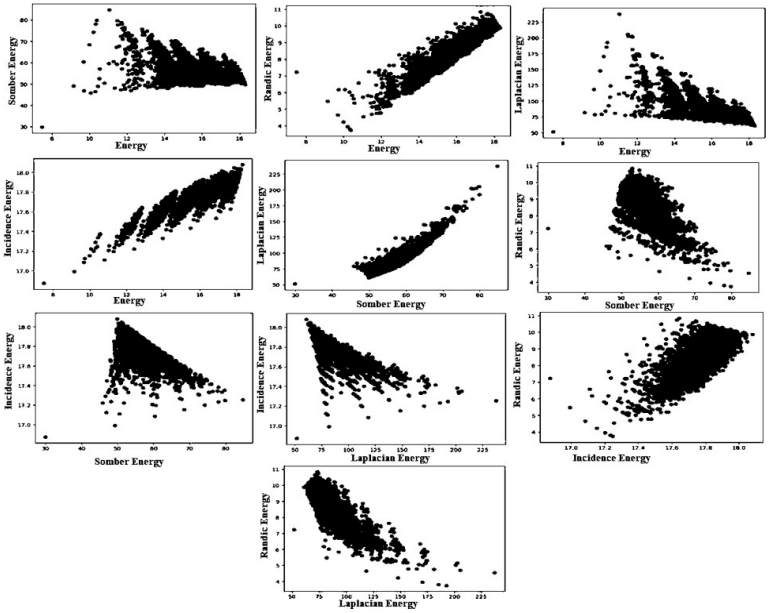

Figure 1. Scatter plots for each pair of energies of 15-vertex trees

Following the selection of the final model by polynomial regression, we investigated the importance of each feature by calculating their im-

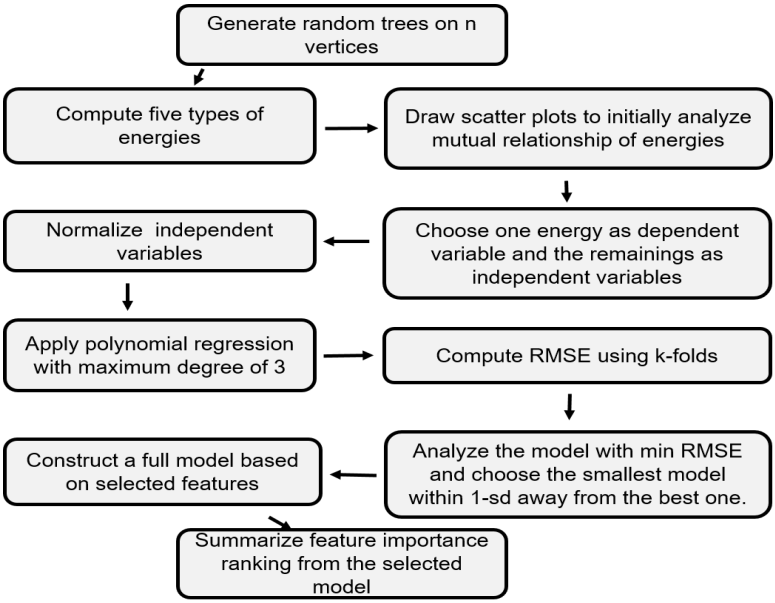

Figure 2. Flowchart depicting the methodology for analyzing relationships between energies in tree graphs

portance. Regardless of directionality, the coefficient associated with each feature measures the strength of the link between the features and the target variable. After taking absolute values, we ranked them in descending order. By examining them, we identified the features with larger magnitudes as having a relatively more substantial influence on the target variable. This ranking provides information about which features contribute the most to the model's predictions.

4 Results and discussion

This section contains the analysis of polynomial regression performance for our considered models. For convenience, we represent ordinary energy, Laplacian energy, Randić energy, incidence energy, and Sombor energy by E, L, R, I, and S respectively. In the tables below, $|\beta|$ is the absolute value of the coefficient of each feature and p-value is from the t-test of

the coefficient given by Equation 1 in preliminaries. The ranking of the features and $p-values < 0.05$ provides insight into which energies are most important for predicting the outcome variable and provides the significance of the relationship between the various energies across the different trees.

Table 2 below presents a series of trees with 6 to 20 vertices representing the relationship of ordinary graph energy with other energies under consideration. Sombor energy has the highest ranking and significantly correlates with ordinary energy for trees on 6 to 9 vertices. With the increasing number of vertices, Randić energy becomes more significant than Sombor energy. From 6 to 9 vertices, the relationship between energies is linear. In contrast, some energies have a quadratic or cubic relationship from 11 to 20 vertices. For example, for trees on 11 vertices, energy is cubically related to Sombor energy. It can also be observed that the RMSE value increases with the increasing number of vertices. For instance, the trees on 6 vertices have an RMSE of 0.01126057, while the trees on 20 vertices yield an RMSE of 0.09582839. This complexity is justified to understand the relationship between these energies better as the number of trees increases exponentially as n increases, as given in Table 1. Overall, our findings suggest that Randić, Sombor, and incidence energies are consistently stronger predictors of ordinary energy.

Table 2 and Table 3 have more complex features for some moderate n. This is due to the randomness of 10-folds that some sensitivity of higher degree terms for moderate n is expected. Even though k-folds involve some randomness, it is still more stable than only splitting data into training and test sets.

Table 3 below presents a series of trees with 6 to 20 vertices representing the relationship of the Laplacian energy with other energies under consideration. From the feature ranking, ordinary energy has the highest ranking. This relationship is linear for 6 to 13 vertices, while it becomes quadratic for 14 to 20 vertices. The second most important feature is Sombor energy, with a linear relationship for all vertices except $n = 11$ which is quadratic. Incidence energy is the third important feature for 6 to 8 vertices while the third important feature for 10 to 20 vertices is Randić energy. For Laplacian energy, the overall most important predictors are

Table 2. Feature ranking for ordinary energy

Vertices		Features Ranking						RMSE
	Name	S	I	L	R			
6	$\|\beta\|$	0.7198	0.3751	0.2711	0.2046			0.01126057
	p-values	0.007	0.004	0.023	0.013			
	Name	S	I	R	L			
7	$\|\beta\|$	0.6961	0.422	0.278	0.269			0.03280668
	p-values	0	0	0	0.004			
	Name	S	I	R	L			
8	$\|\beta\|$	0.6631	0.4487	0.4207	0.296			0.03570834
	p-values	0	0	0	0			
	Name	S	R	I	L			
9	$\|\beta\|$	0.6061	0.5022	0.4661	0.2963			0.04124739
	p-values	0	0	0	0			
	Name	R	S	I	L			
10	$\|\beta\|$	0.6051	0.5793	0.4833	0.3135			0.04269574
	p-values	0	0	0	0			
	Name	R	S	I	L	S^2	S^3	
11	$\|\beta\|$	0.6569	0.5631	0.4826	0.3487	0.0056	0.0021	0.06463317
	p-values	0	0	0	0	0.034	0	
	Name	R	S	I	L	I^2		
12	$\|\beta\|$	0.7248	0.5831	0.5189	0.3515	0.0114		0.06004580
	p-values	0	0	0	0	0		
	Name	R	S	I	L	I^2		
13	$\|\beta\|$	0.7623	0.5944	0.5265	0.3764	0.0122		0.06325146
	p-values	0	0	0	0	0		
	Name	R	S	I	L	I^2		
14	$\|\beta\|$	0.801	0.6157	0.5369	0.4041	0.0128		0.06847567
	p-values	0	0	0	0	0		
	Name	R	S	I	L	I^2		
15	$\|\beta\|$	0.8309	0.6341	0.5454	0.4273	0.0134		0.07340321
	p-values	0	0	0	0	0		
	Name	R	S	I	L	S^2	S^3	
16	$\|\beta\|$	0.8532	0.6525	0.5356	0.4765	0.0054	0.0014	0.08061196
	p-values	0	0	0	0	0	0	
	Name	R	S	I	L	I^2	S^2	
17	$\|\beta\|$	0.8869	0.6766	0.564	0.4764	0.0133	0.002	0.08191707
	p-values	0	0	0	0	0	0	
	Name	R	S	I	L	S^2		
18	$\|\beta\|$	0.9101	0.6912	0.5609	0.5062	0.0111		0.08855195
	p-values	0	0	0	0	0		
	Name	R	S	I	L	S^2		
19	$\|\beta\|$	0.9359	0.7076	0.5721	0.5201	0.0114		0.09272274
	p-values	0	0	0	0	0		
	Name	R	S	I	L	S^2		
20	$\|\beta\|$	0.9615	0.7234	0.5836	0.5326	0.0116		0.09582839
	p-values	0	0	0	0	0		

Note:$|\beta|$ is the absolute value of the coefficient of each feature, and p-value is from the t-test of the coefficient.

ordinary energy and Sombor energy. Trees on 6 vertices have an RMSE of 0.0114757869, while trees on 20 vertices yield an RMSE of 2.67994306. This increase in RMSE is justified as the number of trees increases exponentially as n increases, as given in Table 1.

Table 4 below presents a series of trees with 6 to 20 vertices representing the relationship of the Randić energy with other energies under consideration. From the feature ranking, ordinary energy has the highest ranking. This relationship is linear for 6 to 16 vertices, while for trees on 16 to 20 vertices, this relationship becomes quadratic. The second most important feature is Sombor energy, which has a linear relationship with

Table 3. Feature Ranking for Laplacian Energy

Vertices		Features Ranking					RMSE
	Name	E	S	I	R		
6	$\|\beta\|$	9.9423	8.6005	4.4799	2.4417		0.14757869
	p-values	0.023	0.016	0.024	0.036		
	Name	E	S	I	R		
7	$\|\beta\|$	10.2985	8.5582	4.6807	2.7488		0.52661485
	p-values	0.004	0	0.006	0.033		
	Name	E	S	I	R		
8	$\|\beta\|$	13.7907	10.1367	6.0269	5.1917		0.48064202
	p-values	0	0	0	0		
	Name	E	S	I	R		
9	$\|\beta\|$	15.1972	10.3347	6.5446	6.3363		0.68232998
	p-values	0	0	0	0		
	Name	E	S	R	I	R^2	
10	$\|\beta\|$	18.5661	11.614	9.1922	8.0114	0.4102	0.72626094
	p-values	0	0	0	0	0	
	Name	E	S	R	I	S^2	
11	$\|\beta\|$	19.6049	12.4489	10.1764	8.2866	0.1343	1.5370880
	p-values	0	0	0	0	0.002	
	Name	E	S	R	I		
12	$\|\beta\|$	23.3631	14.2469	13.2798	9.6075		1.5969609
	p-values	0	0	0	0		
	Name	E	S	R	I		
13	$\|\beta\|$	25.0919	15.3421	14.8533	10.1828		1.7908906
	p-values	0	0	0	0		
	Name	E	S	R	I	E^2	
14	$\|\beta\|$	26.0188	16.7166	16.0146	10.9573	0.4997	1.92566508
	p-values	0	0	0	0	0	
	Name	E	S	R	I	E^2	
15	$\|\beta\|$	26.7665	17.5781	16.8033	11.1426	0.5729	2.0336598
	p-values	0	0	0	0	0	
	Name	E	S	R	I	E^2	
16	$\|\beta\|$	27.4506	18.4299	17.4899	11.3007	0.6389	2.1788832
	p-values	0	0	0	0	0	
	Name	E	S	R	I	E^2	
17	$\|\beta\|$	27.9449	19.1443	17.999	11.3881	0.6977	2.2956161
	p-values	0	0	0	0	0	
	Name	E	S	R	I	E^2	
18	$\|\beta\|$	28.4721	19.843	18.4955	11.4972	0.7453	2.4401601
	p-values	0	0	0	0	0	
	Name	E	S	R	I	E^2	
19	$\|\beta\|$	28.9505	20.4751	18.9333	11.5967	0.7843	2.5580593
	p-values	0	0	0	0	0	
	Name	E	S	R	I	E^2	
20	$\|\beta\|$	29.4665	21.0912	19.3766	11.7195	0.814	2.67994306
	p-values	0	0	0	0	0	

Note:$|\beta|$ is the absolute value of the coefficient of each feature, and p-value is from the t-test of the coefficient.

all the vertices. Incidence energy is the third important feature, except trees on 17 vertices. For trees on 11 to 16 vertices, incidence energy has a quadratic relationship with Randić energy. Overall, our findings suggest that ordinary energy and Sombor energy are stronger predictors for Randić energy. Trees on 6 vertices have an RMSE of 0.00139663 and it will increase with the increasing number of vertices as trees on 20 vertices yield an RMSE of 0.02141770.

Table 4. Feature Ranking for Randić Energy

Vertices		Features Ranking					RMSE
	Name	E	S	I	L		
6	$\|\beta\|$	1.4551	1.258	0.6557	0.4736		0.01895269
	p-values	0.013	0.019	0.014	0.036		
	Name	E	S	I	L		
7	$\|\beta\|$	1.3157	0.9688	0.6064	0.3398		0.06502959
	p-values	0	0.001	0	0.033		
	Name	E	S	I	L		
8	$\|\beta\|$	1.3379	0.8604	0.6	0.3544		0.05003650
	p-values	0	0	0	0		
	Name	E	S	I	L		
9	$\|\beta\|$	1.2809	0.7172	0.5726	0.315		0.05173276
	p-values	0	0	0	0		
	Name	E	S	I	L		
10	$\|\beta\|$	1.2849	0.6611	0.571	0.3254		0.05113024
	p-values	0	0	0	0		
	Name	E	S	I	L	I^2	
11	$\|\beta\|$	1.2704	0.6357	0.5781	0.3311	0.0124	0.06430158
	p-values	0	0	0	0	0	
	Name	E	S	I	L	I^2	
12	$\|\beta\|$	1.2921	0.6419	0.5855	0.3617	0.0133	0.06922429
	p-values	0	0	0	0	0	
	Name	E	S	I	L	I^2	
13	$\|\beta\|$	1.304	0.6482	0.5869	0.3865	0.0138	0.07152262
	p-values	0	0	0	0	0	
	Name	E	S	I	L	I^2	
14	$\|\beta\|$	1.3283	0.6674	0.5937	0.4147	0.014	0.07635892
	p-values	0	0	0	0	0	
	Name	E	S	I	L	I^2	
15	$\|\beta\|$	1.3476	0.683	0.5986	0.4367	0.0142	0.08128179
	p-values	0	0	0	0	0	
	Name	E	S	I	L	I^2	
16	$\|\beta\|$	1.3725	0.7016	0.6066	0.4574	0.0143	0.08582208
	p-values	0	0	0	0	0	
	Name	E	S	L	I	E^2	
17	$\|\beta\|$	1.3898	0.7143	0.6055	0.4726	0.0125	0.09219267
	p-values	0	0	0	0	0	
	Name	E	S	I	L	E^2	
18	$\|\beta\|$	1.4131	0.7269	0.6144	0.4821	0.0119	0.09675481
	p-values	0	0	0	0	0	
	Name	E	S	I	L	E^2	
19	$\|\beta\|$	1.4358	0.7382	0.6234	0.4898	0.0112	0.10117061
	p-values	0	0	0	0	0	
	Name	E	S	I	L	E^2	
20	$\|\beta\|$	1.4597	0.7493	0.6331	0.4967	0.0105	0.10442719
	p-values	0	0	0	0	0	

Note:$|\beta|$ is the absolute value of the coefficient of each feature, and p-value is from the t-test of the coefficient.

Table 5 below presents a series of trees on 6 to 20 vertices representing the relationship of the incidence energy with other energies. Ordinary energy has the highest feature ranking for trees on 6 to 20 vertices. This relationship is linear for trees on 6 to 13 vertices and becomes quadratic for trees on 14 vertices, while for trees on 15 to 20 vertices this relationship becomes cubic. The second most important feature is Sombor energy for trees on 6 to 9 vertices while for trees on 10 to 20 vertices, the second most important feature is Randić energy. For $n = 9, 10, 12, 13$, Randić energy has a quadratic relationship with incidence energy. Sombor energy

is the third important feature except for trees on 6 to 9 vertices. Energy and Randić energy are stronger predictors for Incidence energy. Trees on 6 vertices have an RMSE of 0.00139663 and it will increase with the increasing number of vertices as trees on 20 vertices yield an RMSE of 0.02141770.

Table 5. Feature Ranking for Incidence Energy

Vertices		Features Ranking							RMSE
	Name	E	S	L	R				
6	$\|\beta\|$	0.1035	0.0895	0.0337	0.0254				0.00139663
	p-values	0.004	0.009	0.024	0.014				
	Name	E	S	R	L				
7	$\|\beta\|$	0.1328	0.1003	0.0403	0.0385				0.00455069
	p-values	0	0	0	0.006				
	Name	E	S	R	L				
8	$\|\beta\|$	0.1611	0.1055	0.0677	0.0464				0.00574658
	p-values	0	0	0	0				
	Name	E	S	R	L	R^2			
9	$\|\beta\|$	0.1805	0.1047	0.0864	0.0533	0.0033			0.00631308
	p-values	0	0	0	0	0.001			
	Name	E	R	S	L	R^2			
10	$\|\beta\|$	0.1983	0.1087	0.1062	0.0606	0.0032			0.00691777
	p-values	0	0	0	0	0			
	Name	E	R	S	L	L^2			
11	$\|\beta\|$	0.215	0.1247	0.1123	0.0709	0.0019			0.01119425
	p-values	0	0	0	0	0			
	Name	E	R	S	L	R^2			
12	$\|\beta\|$	0.2261	0.1406	0.1146	0.0738	0.0034			0.01285266
	p-values	0	0	0	0				
	Name	E	R	S	L	R^2			
13	$\|\beta\|$	0.2367	0.1522	0.1194	0.0791	0.0035			0.01365461
	p-values	0	0	0	0	0			
	Name	E	R	S	L	E^2			
14	$\|\beta\|$	0.2408	0.1602	0.1239	0.0798	0.003			0.01474971
	p-values	0	0	0	0	0			
	Name	E	R	S	L	E^2	E^3		
15	$\|\beta\|$	0.2494	0.1691	0.1289	0.0839	0.0026	0.0002		0.01596246
	p-values	0	0	0	0	0	0		
	Name	E	R	S	L	E^2	E^3		
16	$\|\beta\|$	0.2581	0.1776	0.1341	0.0877	0.0022	0.0003		0.01719325
	p-values	0	0	0	0	0	0		
	Name	E	R	S	L	E^2	E^3		
17	$\|\beta\|$	0.2659	0.1851	0.1386	0.0907	0.0019	0.0004		0.01833149
	p-values	0	0	0	0	0	0		
	Name	E	R	S	L	E^2	E^3		
18	$\|\beta$	0.2736	0.1923	0.1431	0.0933	0.0016	0.0004		0.01939308
	p-values	0	0	0	0	0	0		
	Name	E	R	S	L	E^2	S^2	E^3	
19	$\|\beta\|$	0.2818	0.199	0.1483	0.0978	0.0013	0.0005	0.0004	0.02048869
	p-values	0	0	0	0	0	0	0	
	Name	E	R	S	L	E^2	S^2	E^3	
20	$\|\beta\|$	0.289	0.2054	0.1522	0.0997	0.0011	0.0005	0.0004	0.02141770
	p-values	0	0	0	0	0	0	0	

Note:$|\beta|$ is the absolute value of the coefficient of each feature, and p-value is from the t-test of the coefficient.

Table 6 below presents a series of trees with 6 to 20 vertices representing the relationship of the Sombor energy with other energies under consideration. From the feature ranking, ordinary energy has the highest ranking and is significantly related to Sombor energy for trees on 6 to 20 vertices.

Table 6. Feature Ranking for Sombor Energy

Vertices		Features Ranking					RMSE
	Name	E	I	L	R		
6	$\|\beta\|$	3.0476	1.3732	0.993	0.7488		0.04242875
	p-values	0.007	0.009	0.016	0.019		
	Name	E	I	L	R		
7	$\|\beta\|$	3.4589	1.5831	1.1108	1.0171		0.10279538
	p-values	0	0	0	0.001		
	Name	E	I	R	L		
8	$\|\beta\|$	4.1247	1.8276	1.6828	1.3532		0.13522622
	p-values	0	0	0	0		
	Name	E	R	I	L		
9	$\|\beta\|$	4.5366	2.105	1.9954	1.5081		0.17879870
	p-values	0	0	0	0		
	Name	E	R	I	L	L^2	
10	$\|\beta\|$	4.8785	2.5397	2.1127	1.9577	0.0607	0.19853543
	p-values	0	0	0	0	0	
	Name	E	R	L	I	L^2	
11	$\|\beta\|$	5.1787	2.8706	2.2423	2.2002	0.0795	0.28483124
	p-values	0	0	0	0	0	
	Name	E	R	I	L	I^2	
12	$\|\beta\|$	5.8869	3.6354	2.6346	2.1584	0.0686	0.32106378
	p-values	0	0	0	0	0	
	Name	E	R	I	L	E^2	
13	$\|\beta\|$	5.9864	3.7938	2.6412	2.4522	0.0899	0.35154649
	p-values	0	0	0	0	0	
	Name	E	R	I	L	E^2	
14	$\|\beta\|$	6.2577	4.0736	2.7419	2.6962	0.0983	0.38178263
	p-values	0	0	0	0	0	
	Name	E	R	L	I	E^2	
15	$\|\beta\|$	6.4439	4.2711	2.8982	2.8079	0.1074	0.41504839
	p-values	0	0	0	0	0	
	Name	E	R	L	I	E^2	
16	$\|\beta\|$	6.6452	4.4683	3.0939	2.8833	0.1143	0.44743566
	p-values	0	0	0	0	0	
	Name	E	R	L	I	E^2	
17	$\|\beta\|$	6.8115	4.6311	3.2589	2.9475	0.1204	0.47508473
	p-values	0	0	0	0	0	
	Name	E	R	L	I	E^2	
18	$\|\beta\|$	6.9881	4.7947	3.412	3.0194	0.1243	0.50443517
	p-values	0	0	0	0	0	
	Name	E	R	L	I	E^2	
19	$\|\beta\|$	7.151	4.9432	3.547	3.0882	0.1272	0.53667942
	p-values	0	0	0	0	0	
	Name	E	R	L	I	E^2	
20	$\|\beta\|$	7.3184	5.0906	3.6729	3.1609	0.1287	0.56269014
	p-values	0	0	0	0	0	

Note:$|\beta|$ is the absolute value of the coefficient of each feature, and p-value is from the t-test of the coefficient.

This relationship is linear for trees on 6 to 12 vertices, while this relationship becomes quadratic with the increasing number of vertices. For trees on 6 to 8, the second important feature is incidence energy; for trees on 9 to 20 vertices, the second important feature is Randić energy. Laplacian energy has a quadratic relationship with Sombor energy for trees on 10 and 11 vertices. Our findings suggest that energy and Randić energies are consistently stronger predictors of Sombor Energy. Trees on 6 vertices have an RMSE of 0.04242875, while trees on 20 vertices yield an RMSE of 0.56269014. This increase in RMSE is justified as the number of trees

increases exponentially as n increases, as given in Table 1.

5 Examples

This section contains three examples of the models produced by the statistical techniques employed. From Table 2 the best model for Laplacian energy for trees on 8 vertices can be given as

$$L = -13.7907(E) + 10.1367(S) + 6.0269(I) + 5.1917(R) + 36.1606 \quad (2)$$

In this model, Laplacian energy is the dependent feature while E, S, R, and I are scaled independent features. As given in Table 1 trees on 8 vertices are 23 in total. We considered the one given in Fig. 3(a) as an example:

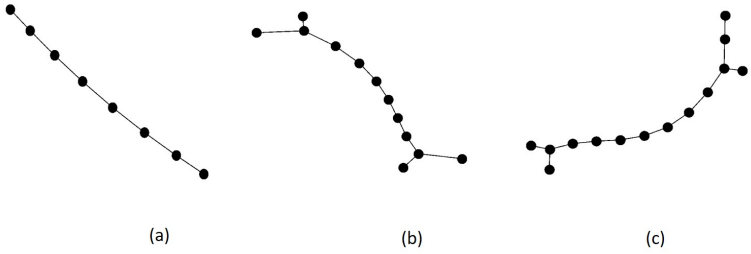

Figure 3. Examples of trees on $n = 8, 12, 14$ vertices

First, we normalize all the independent features. As L is the feature to be predicted in this example so it's not normalized and its actual value is 29.7890. By substituting values of E, S, I and R in Equation 2 it yields,

$$\begin{aligned} L = &-13.7907(1.2551) + 10.1367(-0.4203) + 6.0269(1.5353) \\ &+ 5.1917(1.1687) + 36.1606 = 29.9106 \end{aligned}$$

So, the predicted value of L obtained is 29.9106 which is very close to the actual value.

For our second example, we choose one of the trees on 12 vertices as given in Fig. 3(b), and computations are done to compare actual and

predicted values of Randić energy. The best-chosen model for $n = 12$ is given below:

$$R = 1.2921(E) - 0.6419(S) - 0.5855(I) + 0.3617(L) - 0.0133(I)^2 + 6.9380 \tag{3}$$

Again we normalized independent features i.e. E, S, L and I. While R is in its raw form having an actual value of 7.0288. After substitutions in 3,

$$\begin{aligned} R = {} & 1.2921(0.6509) - 0.6419(-0.7538) - 0.5855(1.4585) \\ & + 0.3617(-0.9025) - 0.0133(1.4585)^2 + 6.9380 = 7.0542 \end{aligned}$$

So, the predicted value of R obtained is 7.0542 which is again very close to the actual value of R.

The next example is for one of the trees on 14 vertices given in 3(c) to analyze model performance for ordinary energy. By using the same method as the previous two examples we normalized independent features and the actual value for the dependent variable i.e. E is 17.0117. Best chosen model for $n = 14$ vertices is given as under

$$E = 0.801(R) + 0.6157(S) + 0.5369(I) - 0.4041(L) + 0.0128(I)^2 + 15.0241 \tag{4}$$

After required substitutions in Equation 4.

$$\begin{aligned} E = {} & 0.801(0.2992) + 0.6157(-1.0895) + 0.5369(1.8012) \\ & - 0.4041(-1.1202) + 0.0128(1.8012)^2 + 15.0241 = 16.0543. \end{aligned}$$

The predicted value of E obtained is 16.0543 which is close to the actual value of E.

6 Concluding remarks

From the results above, the ordinary energy of the graph is the first important feature to predict all other energies i.e. Laplacian, Randić, incidence, and Somber. It may be due to the fact that the concept of simple energy

serves as the foundational definition of energy. It is from this basic understanding of energy that all other forms of energy are related to ordinary energy. Of all the energies considered, incidence and Laplacian energies are defined in different manners. For example, the incidence energy is defined based on an incidence matrix which is not a square matrix that involves the relationship between the graph's vertices and edges and multiplied by its transpose first to get the singular values. While the Laplacian energy involves the average degree of the graph along with the eigenvalues of the Laplacian matrix. Due to their distinct definitions, these two energy measures are ranked third or fourth in the feature ranking. RMSE values for predicting ordinary, Randić, and incidence energies are pretty good, as the RMSE for ordinary energy lies between 0.01126057 and 0.09582839, and Randić energy between 0.01895269 and 0.10442719. For incidence energy, it lies between 0.00139663 and 0.021417. For Laplacian and Sombor energy RMSE is not as small as other energies. For Laplacian energy, RMSE lies between 0.14757869 and 2.67994306, and for Somber energy, it lies between 0.04242875 and 0.56269014. Still, it is acceptable as the values of Somber and Laplacian energies are significantly larger than the other three energies.

***Acknowledgment*:** Hafsah Tabassum has been supported by Petchra Pra Jom Klao Ph.D. Research Scholarship from King Mongkut's University of Technology Thonburi (1281/2021). Pawaton Kaemawichanurat has been supported by National Research Council of Thailand (NRCT) and King Mongkut's University of Technology Thonburi (N42A660926).

References

[1] A. S. Angadi , M. Hatture. Face recognition through symbolic modeling of face graphs and texture. *Int. J. Pattern Recogn. Art. Intell.* **33** (2019) #1956008.

[2] M. Akram, S. Naz, Energy of Pythagorean fuzzy graphs with applications,*Mathematics* **6** (2018) #136.

[3] A. T. Balaban, D. Ciubotariu, M. Medeleanu, Topological indices and real number vertex invariants based on graph eigenvalues or eigenvectors, *J. Chem. Inf. Comput. Sci.* **31** (1991) 517–523.

[4] S. B. Bozkurt, A. D. Güngör, I. Gutman, Randić matrix and Randić energy, *MATCH Commum. Math. Comput. Chem.* **64** (2010) 239–250.

[5] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory*, Princeton Univ. Press, Princeton, 2005.

[6] L. Di Paola, G. Mei, A. Di Venere, A. Giuliani, Exploring the stability of dimers through protein structure topology, *Curr. Protein Peptide Sci.* **17** (2016) 30–36.

[7] D. Cvetković, P. Rowlinson, S. K. Simić, *An Introduction to the Theory of Graph Spectra*, Cambridge Univ. Press, Cambridge, 2010.

[8] D. Cvetković, M. Doob, H. Sachs, *Spectra of Graphs – Theory and Application*, Academic Press, New York, 1980.

[9] K. C. Das, I. Gutman, I. Milovanović, B. Furtula, Degree-based energies of graphs, *Lin. Algebra Appl.* **554** (2018) 185–204.

[10] I. Gutman, The energy of a graph, *Ber. Math. Stat. Sekt. Forschungsz. Graz* **103** (1978) 1–22.

[11] I. Gutman, Total π-electron energy of benzenoid hydrocarbons, *Topics Curr. Chem.* **162** (1992) 29–63.

[12] I. Gutman, The energy of a graph: Old and new results, in: A. Betten, A. Kohnert, R. Laue, A. Wassermann (Eds.), *Algebraic Combinatorics and Applications*, Springer, Berlin, 2001, pp. 196–211.

[13] I. Gutman, Topology, and stability of conjugated hydrocarbons. The dependence of total π-electron energy on molecular topology, *J. Serb. Chem. Soc.* **70** (2005) 441–456.

[14] I. Gutman, B. Furtula, Graph energies and their applications, *Bull. Sci. Math.* **44** (2019) 29–45.

[15] J. Jiang, R. Zhang, L. Guo, W. Li, X. Cai, Network aggregation process in multilayer air transportation networks, *Chin. Phys. Lett.* **33** (2016) #108901.

[16] X. Li, Z. Wang, Trees with the extremal spectral radius of weighted adjacency matrices among trees weighted by degree-based indices, *Lin. Algebra Appl.* **620** (2021) 61–75.

[17] L. Lovász, J. Pelikán, On the eigenvalues of trees, *Period. Math. Hung.* **3** (1973) 175–182.

[18] L. Magee, Non-local behavior in polynomial regressions, *Am. Stat.* **52** (1998) 20–22.

[19] B. J. McCelland, Properties of the latent roots of a matrix: The estimation of π-electron energies, *J. Chem. Phys.* **54** (1971) 640–643.

[20] M. Morzy, T. Kajdanowicz, Graph energies of egocentric networks and their correlation with vertex centrality measures, *Entropy* **20** (2018) #916.

[21] Y. Shao, Y. Gao, W. Gao, X. Zhao, Degree-based energies of trees, *Lin. Algebra Appl.* **621** (2021) 18–28.

[22] D. Stevanović, I. Stanković, M. Milošević, More on the relation between energy and Laplacian energy of graphs, *MATCH Commun. Math. Comput. Chem.* **61** (2009) 395–401.

[23] R. Tibshirani, T. Hastie, J. H. Friedman, *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*, Springer, New York, 2001.

[24] P. Van Mieghem, *Graph Spectra for Complex Networks*, Cambridge Univ. Press, Cambridge, 2010.

[25] K. Yuge, Graph representation for configuration properties of crystalline solids, *J. Phys. Soc. Japan* **86** (2017) #024802.