

WIENER INDEX ESTIMATIONS OF
SECOND-LARGEST EIGENVALUES ACROSS
CUBIC GRAPH ORDERS

BY

SHAHAD ABDULAZIZ ALDOSSARY

A Thesis Presented to the
COLLEGE OF GRADUATE AND INTERDISCIPLINARY STUDIES
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS

DHAHRAN, SAUDI ARABIA

In Partial Fulfillment of the
Requirements for the Degree of

MASTER OF SCIENCE
In
MATHEMATICS

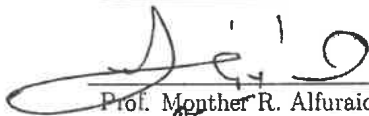
MAY, 2024

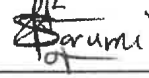
KING FAHD UNIVERSITY OF PETROLEUM & MINERALS
DHAHRAN 31261, SAUDI ARABIA

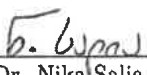
DEANSHIP OF GRADUATE STUDIES

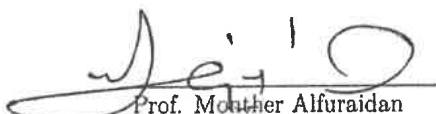
This thesis, written by **SHAHAD ABDULAZIZ ALDOSSARY** under the direction of her thesis adviser and approved by her thesis committee, has been presented to and accepted by the Dean of Graduate Studies, in partial fulfillment of the requirements for the degree of **MASTER OF SCIENCE IN MATHEMATICS**.

Thesis Committee


Prof. Monther R. Alfuraidan (Adviser)


Dr. Ibrahim O. Sarumi (Member)


Dr. Nika Salia (Member)


Prof. Monther Alfuraidan
Department Chairman


Prof. Suliman Al-Homidan
Dean of Graduate Studies

Date



©Shahad A. Aldossary
2024

To my parents

ACKNOWLEDGMENTS

First and foremost, I would like to express my sincere gratitude to Allah, the Most Gracious and Most Merciful, for the endless blessings He has showered upon me in the process of completing this thesis. I am beyond grateful for His blessings, which have broadened my understanding and strengthened my will to advance knowledge. All praise to Allah, the Giver of all wisdom and knowledge.

Then I extend my deepest appreciation and gratitude to Prof. Monther Alfuraidan for all of his assistance, guidance, and encouragement throughout the course of my thesis. His expertise and supervision have been very helpful in guiding my study and polishing my academic skills. I am truly grateful for his patience and encouragement, which have motivated me to pursue academic goals and strive for excellence. I am honored to have had the privilege of working under his supervision and am thankful for the knowledge and experiences gained under his mentorship.

Also I am deeply grateful to Dr. Ibrahim Sarumi and Dr. Nika Salia for their invaluable contributions, support, and expertise as members of my thesis committee. Their insightful comments, constructive criticisms, and support have greatly enhanced my research and academic journey. I extend a special acknowledgment to Dr. Ibrahim Sarumi for his exceptional guidance, mentorship, and unwavering commitment to

excellence. His expertise, dedication, and passion for academia have been a source of inspiration throughout this endeavor. I am privileged to have had the opportunity to learn from their wealth of knowledge, and I am truly thankful to their invaluable contributions to my academic growth.

I extend my genuine gratitude to Prof. Othman Echi for his exceptional teaching and guidance in the fields of graph theory and combinatorics. I am deeply grateful for his insightful lectures, challenging assignments, and encouragement, which have inspired me to delve deeper into the complexities of graph theory and combinatorics. I am honored to have had the opportunity to learn from such a knowledgeable and dedicated professor, and I sincerely value the contributions he has made to my advancement as a scholar.

I am profoundly grateful to my parents for the love, encouragement, and sacrifices throughout this journey. Their endless support, and belief in my abilities have been the driving force behind my success. I owe them a debt of gratitude for helping me develop the qualities of perseverance, determination, and hard work that have molded me into the person I am today. Only Allah can help me reward my parents, the most precious blessing of Allah. Additionally, I would like to acknowledge my siblings for their support, and understanding. Their encouragement and belief in my potential have been a source of strength and inspiration during challenging times. I am blessed to have such a loving and supportive family, and I am deeply thankful for their unwavering support throughout my academic endeavors.

Contents

ACKNOWLEDGMENTS	v
LIST OF TABLES	ix
LIST OF FIGURES	x
ABSTRACT (ENGLISH)	xiii
ABSTRACT (ARABIC)	xv
CHAPTER 1 INTRODUCTION	1
1.1 Structure of the Thesis	1
CHAPTER 2 PRELIMINARIES	3
2.1 Fundamental Concepts of Graph Theory	3
2.2 Graph Theory Applications in Chemistry: Understanding Topological Indices	8
2.2.1 Other Notable Topological Indices	11
2.2.2 Significance of Distance-Based Indices in Regular Graphs	13
CHAPTER 3 ADVANCED INSIGHTS IN REGRESSION ANAL- YSIS: RELATIONSHIPS AND PREDICTIVE MODELING	18
3.1 Linear Regression Analysis	20
3.1.1 Simple Linear Regression	20
3.1.2 Multiple Linear Regression	23
3.1.3 Method of Least Squares	25
3.1.4 Method of Maximum Likelihood	28

3.1.5	Method of Maximum Likelihood for Multiple Linear Regression	30
3.1.6	Normal Equation for Linear Regression Models	32
3.2	Coefficient of Determination	35
CHAPTER 4 REFINED MODELING AND DATA ANALYSIS		
TECHNIQUES		38
4.1	Data Extraction	38
4.2	Development of the Regression Model	39
4.2.1	Data Visualization	39
4.2.2	Exploration of Regression Models for Cubic Graphs of Orders 10 to 20	41
4.2.3	Model Validation for Cubic Graphs of Order 10	47
4.2.4	Model Validation for Cubic Graphs of Orders 12, 14, 16, 18, 20	51
4.2.5	Practical Application	57
4.2.6	Average of the Wiener Index	59
CHAPTER 5 MODEL GENERALIZATION FOR CUBIC GRAPHS		66
5.1	Development of a Generalized Regression Model	66
5.2	Enhanced Predictive Modeling Using Wiener Index Averages	69
5.2.1	Exploring Wiener Indices Below the Average	71
CHAPTER 6 CONCLUSION AND RECOMMENDATIONS		75
6.1	Conclusion	75
6.2	Recommendations for Future Work	76
REFERENCES		78
VITAE		80

List of Tables

Table 2.1	Molecular diagrams of n - butane and n - pentane.	9
Table 2.2	Skeletal diagrams of n - butane and n - pentane.	10
Table 2.3	Closed forms of Wiener indices for some classical graphs	14
Table 2.4	The number of connected cubic graphs on a given number of vertices, illustrating the exponential growth in complexity with increasing vertices, see [1].	16
Table 3.1	Data from the Persistence Study	26
Table 3.2	Data from the Height and Weight Study	29
Table 4.1	Comparing the R-Squared for different models of G_{10}	42
Table 4.2	Comparing the R-Squared for different models of G_{12}	43
Table 4.3	Comparing the R-Squared for different models of G_{14}	44
Table 4.4	Comparing the R-Squared for different models of G_{16}	45
Table 4.5	Comparing the R-Squared for different models of G_{18}	46
Table 4.6	Comparing the R-Squared for different models of G_{20}	47

List of Figures

Figure 2.1	A connected simple graph of order 5 and size 7	4
Figure 2.2	Examples of 1-regular, 2-regular, and 3-regular graphs, showcasing varying degrees of regularity.	5
Figure 2.3	Visual representation of graph isomorphism between two graphs. . .	5
Figure 2.4	Nonisomorphic graphs demonstrating how structural differences im- pact connectivity and graph properties.	6
Figure 2.5	Wiener’s resulting table of Normal Paraffins	10
Figure 2.6	Different structures of cubic graphs with 12 vertices	15
Figure 3.1	Visualization of the Least Squares Criterion Q for Regression Line Fit - Example from the Persistence Study.	27
Figure 3.2	A scatter plot with original points and a regression line (orange). . .	35
Figure 3.3	Comparison of Two Linear Regression Models	36
Figure 3.4	A model with high R - squared	37
Figure 3.5	A model with low R - squared	37
Figure 4.1	Scatter plots illustrating the relationship between the Wiener in- dices of cubic graphs and λ_2 across different orders.	40
Figure 4.2	Correlation between the Wiener index and the estimated second- largest eigenvalue λ_2 for cubic graphs of order 10.	50
Figure 4.3	Analysis of model performance based on training and testing data for cubic graphs of order 10.	50
Figure 4.4	Exploration of the connection between Wiener’s index and λ_2 of cubic graphs of order 12	53

Figure 4.5	Indications derived from the training and testing datasets for order 12.	54
Figure 4.6	Models from the parameters obtained from the corresponding dataset for orders 14, 16, 18, and 20.	55
Figure 4.7	Predictions of the training and testing data for orders 14 and 16. . .	56
Figure 4.8	Predictions of the training and testing data for orders 18 and 20. . .	57
Figure 4.9	Exploration of the connection between Wiener's index and λ_2 for cubic graphs of order 12 with Wiener indices $\geq$ the average. . . .	60
Figure 4.10	Model performance and error distribution for cubic graphs of order 12. These plots are critical for evaluating the accuracy of the regression model and for understanding the variability of the model's predictions compared to actual data.	61
Figure 4.11	The percentage error of the testing data for different cases. This figure breaks down the error percentage based on whether the Wiener index is greater than, less than, or equal to the average, providing deeper insight into model performance across these subgroups. . . .	61
Figure 4.12	Exploration of the connection between Wiener's index and λ_2 of cubic graphs of order 14, for the dataset with $W(G_{14}) \geq 213$	62
Figure 4.13	The percentage error of the testing data for cubic graphs of order 14, highlighting precision in model predictions.	62
Figure 4.14	The relation between the data points where $W(G_{16}) \geq 299$ and λ_2 , illustrating the model's accuracy for high Wiener indices.	63
Figure 4.15	Exploration of the connection between Wiener's index and λ_2 across different cubic graph orders, confirming consistency and reliability of the predictive model.	63
Figure 5.1	Scatter plot exploring the relationship between the Wiener index and λ_2 across cubic graphs of various orders	67
Figure 5.2	Analysis of prediction errors for the second-largest eigenvalue in graphs of order 14	68

Figure 5.3	Evaluation of the generalized regression model's performance for orders 16 and 18	68
Figure 5.4	Analysis of the relationship between Wiener's index and λ_2 for cubic graphs G of various orders n , where $W(G) \geq \mu_n$	69
Figure 5.5	Evaluation of percentage error in predicting the second-largest eigenvalue for graphs of order 14	70
Figure 5.6	Assessment of the percentage error in estimating the second-largest eigenvalue for graphs of orders 16 and 18	71
Figure 5.7	Analysis of percentage error in estimating the second-largest eigenvalue for graphs of order 14 with Wiener indices below the average .	72
Figure 5.8	Evaluation of percentage errors in estimating the second-largest eigenvalues for graphs with 16 and 18 vertices	73
Figure 5.9	The percentage error for estimating the second-largest eigenvalue for graphs with 20 vertices using a model focused on graphs with Wiener indices below the average	74

THESIS ABSTRACT

NAME: Shahad Abdulaziz Aldossary

TITLE OF STUDY: Wiener Index Estimations of Second-Largest Eigenvalues
Across Cubic Graph Orders

MAJOR FIELD: Mathematics

DATE OF DEGREE: May, 2024

This thesis explores connected cubic graphs, with a specific aim to approximate the second-largest eigenvalue using the Wiener index. We begin with an introduction to the foundational concepts of graphs, matrices, and eigenvalues, and highlight significant results on regular graphs. The study also delves into various notable topological indices, focusing on those based on vertex degrees and distances. Regression analysis techniques are introduced, and simple examples are used to demonstrate their practical application. The thesis briefly discusses the coefficient of determination, underscoring its importance in evaluating the effectiveness of a regression model. This study proposes a model based on the Wiener index for cubic graphs of orders up to 20 to approximate the second-largest eigenvalues, which are challenging to calculate from

adjacency matrices, especially for graphs with a large number of vertices. The model aims to elucidate the relationship between graph properties and connectivity metrics.

ملخص الرسالة

الاسم: شهد عبدالعزيز الدوسري

عنوان الدراسة: التقريبات للقيمة الثانية الأكبر للقيم الذاتية للرسوم المنتظمة الثلاثية

التخصص: قسم الرياضيات

تاريخ الدرجة العلمية: مايو، 2024

تستكشف هذه الرسالة الرسوم البيانية المكعبة المتصلة، بهدف تقريب ثاني أكبر قيمة ذاتية λ_2 باستخدام مؤشر وينر. نبدأ بمقدمة للمفاهيم الأساسية للرسوم البيانية، المصفوفات، والقيم الذاتية، ونسلط الضوء على النتائج الهامة المتعلقة بالرسوم المنتظمة. تتناول الدراسة أيضًا مؤشرات توبولوجية بارزة مختلفة، مع التركيز على تلك التي تستند إلى درجات الرؤوس والمسافات. يتم تقديم تقنيات تحليل الانحدار، ويتم استخدام أمثلة بسيطة لتوضيح تطبيقها العملي. تناقش الرسالة بإيجاز معامل التحديد، مؤكدة أهميته في تقييم فعالية نموذج الانحدار. تقترح هذه الدراسة نموذجًا يستند إلى مؤشر وينر للرسوم المكعبة ذات الطلبات حتى 20 لتقريب ثاني أكبر القيم الذاتية، والتي يصعب حسابها من مصفوفات التجاور، خاصةً للرسوم ذات عدد كبير من الرؤوس. يهدف النموذج إلى توضيح العلاقة بين خصائص الرسوم ومقاييس الاتصال.

CHAPTER 1

INTRODUCTION

Graph theory has emerged as an indispensable tool in modern mathematics, influencing diverse fields such as computer science, engineering, and biology through its exploration of vertices and edges in various configurations. This thesis focuses on a specific aspect of graph theory: the approximation of the second-largest eigenvalues of cubic graphs using the Wiener Index. The Wiener Index, a well-established topological index, is employed due to its strong correlation with several properties of graphs that are important in theoretical and applied settings.

1.1 Structure of the Thesis

This thesis is meticulously organized to provide a comprehensive understanding of how topological indices, particularly the Wiener Index, can be applied to estimate important spectral properties of cubic graphs. The organization is as follows:

- **Chapter 2: Preliminaries** - This chapter lays the foundational knowledge required for understanding the complex theories discussed in later chapters. It

covers fundamental concepts of graph theory and explores the application of these concepts in chemistry, focusing on various topological indices and their significance in regular graphs.

- **Chapter 3: Advanced Insights in Regression Analysis: Relationships and Predictive Modeling** - Delving deeper into statistical methodologies, this chapter explores various regression analysis techniques, emphasizing their application in predicting and modeling graph properties.
- **Chapter 4: Refined Modeling and Data Analysis Techniques** - Focused on practical application, this chapter details the data extraction processes, development, and validation of regression models for cubic graphs of various orders, using the Wiener Index as a predictive tool.
- **Chapter 5: Model Generalization for Cubic Graphs** - Extending the application of developed models, this chapter discusses the generalization of these models across different cubic graphs, enhancing their predictive power with the aid of the Wiener Index.
- **Chapter 6: Conclusion and Recommendations** - The final chapter summarizes the findings and discusses the implications of the research, offering recommendations for future work to extend and refine the methodologies presented.

Through the structured exploration of these topics, the thesis aims to enhance the understanding and application of topological indices in graph theory, particularly the Wiener Index, in estimating critical spectral properties of cubic graphs.

CHAPTER 2

PRELIMINARIES

2.1 Fundamental Concepts of Graph Theory

This section provides a comprehensive review of the fundamental concepts and terminology essential for understanding the core aspects of this thesis related to graph theory.

Definition 2.1 *A **graph** G comprises three elements: a set of vertices written as $V(G)$, a set of edges represented by $E(G)$, and a relationship that links every edge with two not necessarily distinct vertices, termed as its endpoints. The **size of** G is represented by m , which equals the cardinality of the edge set $|E(G)|$. The vertex set's total quantity of vertices $|V(G)|$ is the **order of** G , represented by n . For instance, a graph with vertices $\{5, 1, 4, 3, 2\}$ and edges $\{12, 23, 34, 45, 51, 13, 24\}$ is a **connected simple graph**.*

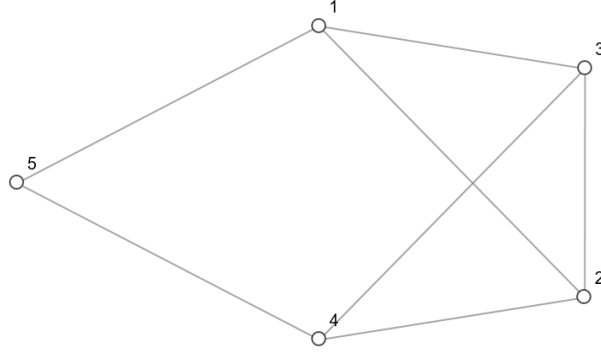


Figure 2.1: A connected simple graph of order 5 and size 7

Definition 2.2 The term **connected** refers to a graph G where a path linking every two vertices exists; if not, G is defined as **disconnected** since there are vertices which have no path joining them.

Definition 2.3 A **loop** is defined as an edge with equal endpoints. Edges that have the same endpoints are referred to as **Multiple edges**. A graph is referred to as **simple** if it has no loops or multiple edges.

In this study, the focus is primarily on simple connected graphs. We begin by defining basic terms and proceed to explore significant matrix representations such as the adjacency matrix and the graph spectrum.

Definition 2.4 v and e are incident if edge e has a vertex v as its endpoint. The total number of edges that meet with v , indicated by $d(v)$, is its **degree**. $\delta(G)$ represents the graph's minimum degree of vertices, while $\Delta(G)$ represents the (finite) graph's highest degree of vertices.

If the degree of each vertex v in V is k , then the graph G is **k -regular**. A graph is said to be **regular** if it is k -regular for some k . This leads to the special case of

cubic graphs, which are 3-regular graphs, forming an important part of this study.

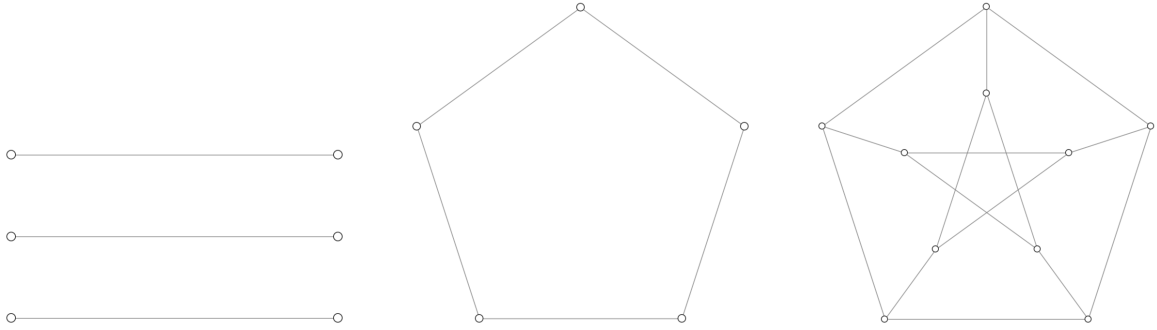


Figure 2.2: Examples of 1-regular, 2-regular, and 3-regular graphs, showcasing varying degrees of regularity.

Definition 2.5 Two graphs G and H are **isomorphic**, denoted as $G \cong H$, if a bijection exists between their vertex sets while preserving adjacency.

The notion of isomorphism implies that isomorphic graphs have identical graph spectra and therefore, share many structural properties, which makes it a fundamental aspect in the study of graph invariants.

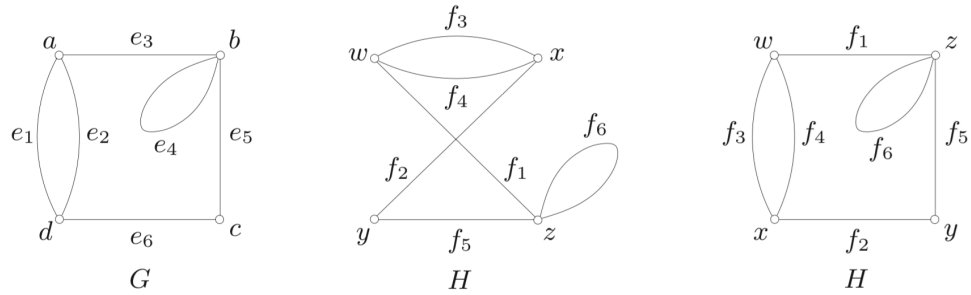


Figure 2.3: Visual representation of graph isomorphism between two graphs.

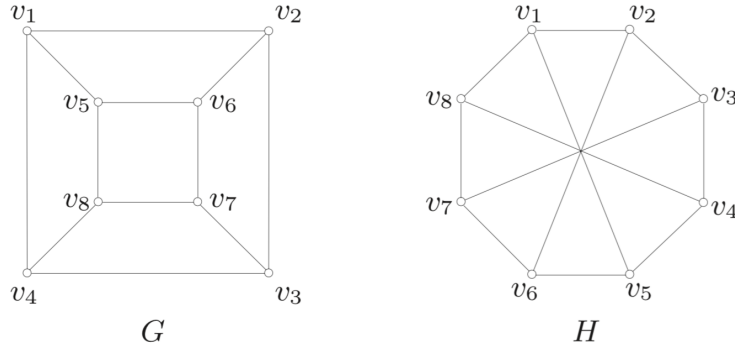


Figure 2.4: Nonisomorphic graphs demonstrating how structural differences impact connectivity and graph properties.

Definition 2.6 A graph G having n vertices has an **adjacency matrix** $A(G)$, which is a square matrix of size n having elements a_{ij} , which gives the entry 1 when vertex i is adjacent to vertex j and 0 otherwise. That is,

$$a_{ij} = \begin{cases} 1, & \text{if } i \sim j \\ 0, & \text{otherwise.} \end{cases}$$

The adjacency matrix is fundamental in understanding the structure of a graph as it encodes the connectivity between vertices. It is particularly essential in calculating the graph's spectrum, which is crucial for the analysis of graph properties.

Definition 2.7 The **Laplacian matrix** L of a graph G is defined as $L = D - A$, where D is the degree matrix — a diagonal matrix where each element d_{ii} is the degree of vertex i , and A is the adjacency matrix of G .

The Laplacian matrix plays a key role in graph theory, particularly in the study of graph connectivity and the calculation of network robustness. It is also used in various

applications such as physics, chemistry, and engineering to solve problems related to network dynamics.

Definition 2.8 *The **eigenvalues** of a graph are derived from the adjacency or Laplacian matrix. The spectrum of a graph, which includes all its eigenvalues, provides valuable insights into the graph's properties such as connectivity and stability.*

The spectrum of a graph is crucial for understanding its overall structure and behavior, particularly in complex networks. The eigenvalues can indicate the presence of community structures, the robustness of the network to attacks, and many other important characteristics.

These fundamental matrix concepts form the basis for the advanced topics discussed in this thesis, including the use of the Wiener Index for estimating the second-largest eigenvalues of cubic graphs. As we explore the relation between the eigenvalues and the Wiener Index, these matrix representations provide the necessary theoretical foundation for our analysis.

As this thesis progresses, these basic concepts from graph theory will serve as the backbone for deeper analyses into the properties of cubic graphs, particularly in exploring the relationship between their topological indices and the eigenvalues derived from their adjacency matrices.

2.2 Graph Theory Applications in Chemistry: Understanding Topological Indices

Chemists often envision molecules as geometric structures composed of atoms in specific configurations. Despite constant intra-molecular movements, the fundamental structure of a molecule remains intact due to certain invariant properties. This phenomenon led to the development of topology, a branch of mathematics focusing on properties that are preserved under continuous deformations.

In chemical graph theory, molecular structures are represented as graphs, with vertices representing atoms and edges representing chemical bonds. Although the topological representation of a molecule is inherently qualitative, numerical descriptors called topological indices quantify these structures, aiding in the prediction of molecular properties in complex experiments.

Research such as [2], [3] highlights the importance of molecular graphs in computational chemistry and theoretical studies. Extensive studies in [4], [5], [6], [7] have explored various topological indices, revealing their potential in understanding molecular behaviors.

An important topological index, known as the ***Wiener index***, was proposed by Harry Wiener in 1947 [8]. It was initially designed for characterizing alkane molecules, it stands as one of the earliest recognized topological indices. The Wiener index is essentially the total of the shortest paths connecting every pair of vertices within a graph and has been widely used to predict molecular properties such as boiling points.

Definition 2.9 For a simple connected graph G , the Wiener index, denoted by $W(G)$, is defined as:

$$W(G) = \sum_{1 \leq i < j \leq n} d_G(v_i, v_j)$$

where $d_G(v_i, v_j)$ is the shortest path length between vertices v_i and v_j .

The predictive power of the Wiener index was demonstrated by its ability to estimate the boiling points of paraffins, a class of hydrocarbons with the general formula C_nH_{2n+2} . This correlation is encapsulated in the equation:

$$\Delta t = (98/n^2)\Delta w + 5.5\Delta p$$

where Δw and Δp are structural variables reflecting changes in molecular geometry.

To clarify the relationship between Wiener's index and the boiling points of paraffins, consider the following representation. Initially, molecular diagrams of paraffins such as n-butane and n-pentane are abstracted into their skeletal forms. These skeletal representations help in computing the path number, known as the Wiener index, for these compounds.

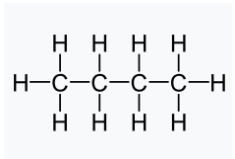
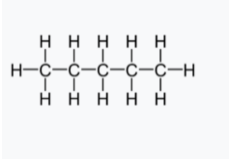
n-butane	n-pentane
	

Table 2.1: Molecular diagrams of n -butane and n -pentane.

The skeletal diagrams used to calculate the Wiener index for these molecules are

shown in the Figure below.

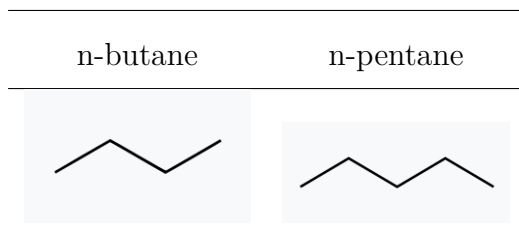


Table 2.2: Skeletal diagrams of n - butane and n - pentane.

After computing the path number for the paraffins from their skeletal diagrams, Harry Wiener formulated a relationship between these indices and the boiling points of the molecules. This relationship is quantitatively demonstrated in the following table:

TABLE I			
NORMAL PARAFFINS			
Cpd.	t_b	w_0	p_0
<i>n</i> -Butane	-0.5	10	1
<i>n</i> -Pentane	36.1	20	2
<i>n</i> -Hexane	68.7	35	3
<i>n</i> -Heptane	98.4	56	4
<i>n</i> -Octane	125.7	84	5
<i>n</i> -Nonane	150.8	120	6
<i>n</i> -Decane	174.0	165	7
<i>n</i> -Undecane	195.8	220	8
<i>n</i> -Dodecane	216.2	286	9

Figure 2.5: Wiener's resulting table of Normal Paraffins

Harry Wiener concluded that the boiling point decreases as the path number w decreases. Additionally, it is observed that the boiling point varies inversely with the compactness of the molecular structure. This inverse relationship highlights that molecules with more compact structures, indicated by a lower Wiener index, generally possess higher boiling points compared to less compact structures.

2.2.1 Other Notable Topological Indices

Here, we explore several other significant topological indices and provide their mathematical formulations, which have proven crucial in characterizing various molecular properties.

Hosoya Index (Z Index): Proposed by Haruo Hosoya in 1971 [9], it is defined as the total number of matchings in a graph, and sometimes termed as the Z index. It reflects the number of ways edges can be chosen where no two edges share a common vertex. The index is mathematically formulated as:

$$Z(G) = \sum_{k=0}^n m_k(G)$$

where $m_k(G)$ represents the count of k -matchings in the graph G .

Zagreb Indices: Developed by Croatian chemists in 1972, the Zagreb indices come in two forms: - The calculation of *the first Zagreb index* (M_1) involves summing the vertices' degrees in the molecular graph raised to the second power, expressed mathematically as:

$$M_1 = M_1(G) = \sum_{v \in V(G)} d^2(v),$$

- The second Zagreb index (M_2) is computed as the total of the products of the degrees of pairs of adjacent vertices, that is:

$$M_2 = M_2(G) = \sum_{uv \in E(G)} d(u)d(v).$$

Randic Index: Proposed by Milan Randic in 1975 [10], this index is a measure

of the molecule’s branching and is calculated using:

$$R(G) = \sum_{uv \in E(G)} \frac{1}{\sqrt{d(u) \cdot d(v)}}$$

where $d(u)$ and $d(v)$ represent the degrees of vertices u and v , respectively, that share an edge.

Balaban Index (J Index): This index, proposed by Alexandru T. Balaban in 1982 [11], [12], combines both distance and degree properties:

$$J(G) = \frac{1}{m - n + 2} \sum_{uv \in E(G)} \frac{1}{\sqrt{d(u) + d(v) - 2}}$$

where m and n denote the edge and vertex counts of graph G , respectively.

Atom-Bond Connectivity (ABC) Index: Introduced in the late 1990s, the ABC index evaluates the ease of electron delocalization in a molecule and is given by:

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

where uv represents an edge connecting vertices u and v , and $d(u)$ and $d(v)$ represent the degrees of these respective vertices.

These indices are not only instrumental in the development of quantitative structure-property relationships (QSPR) and structure-activity relationships (QSAR) but are crucial for the discovery and development of new chemical entities. Each index provides unique insights into the molecular framework and its chemical behavior, offering quantitative measurements that help predict properties such as boiling points,

stability, and reactivity based on the molecular structure.

2.2.2 Significance of Distance-Based Indices in Regular Graphs

In the context of regular graphs, where each vertex has the same degree, degree-based indices become less informative. This limitation highlights the significance of distance-based indices such as the Wiener index, which provide more nuanced insights into graph structure and connectivity. To illustrate the practical application and relevance of the Wiener index in graph theory and chemistry, we consider the following Proposition:

Proposition 1 *Consider a regular graph G with degree k . It is fundamental to understand that:*

- i. k itself represents the first largest eigenvalue of G ;*
- ii. if G is connected, then this eigenvalue k has a multiplicity of 1;*

This proposition, while basic, provides a direct view of the first largest eigenvalues and is crucial for setting the stage for deeper analysis.

To offer a general idea about the Wiener index values across various well-known classes of graphs, we include the following table:

Graph G	$W(G)$
Complete graph K_n	$\frac{1}{2}n(n-1)$
Cycle graph $C_n, n \geq 3$	$\begin{cases} \frac{n^3}{8} & \text{for } n \text{ even} \\ \frac{1}{8}n(n^2-1) & \text{for } n \text{ odd} \end{cases}$
Path graph P_n	$\frac{1}{6}n(n^2-1)$
Star graph S_n	$(n-1)^2$
Wheel graph W_n	$(n-2)(n-1)$

Table 2.3: Closed forms of Wiener indices for some classical graphs

Figure 2.6 below is included to highlight that despite identical largest eigenvalues, graphs can exhibit varying levels of compactness, a property not adequately captured by simply considering the largest eigenvalue λ_1 . This observation underscores the inadequacy of λ_1 in fully characterizing a graph's structure, particularly in complex or irregular topologies:

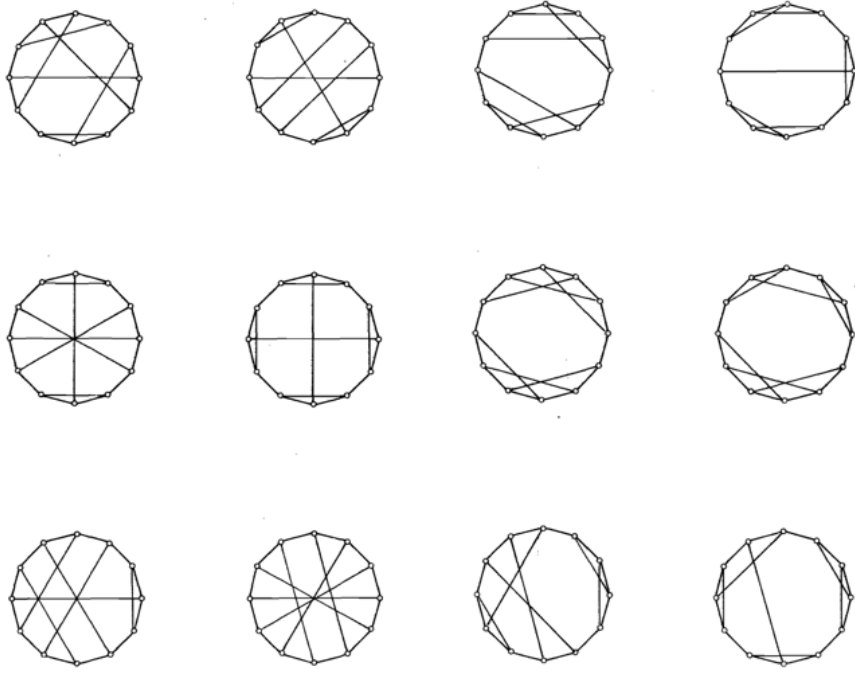


Figure 2.6: Different structures of cubic graphs with 12 vertices

Furthermore, given the complexity and the difficulty in calculating eigenvalues for large or irregular graphs, the Wiener index offers a simpler, yet effective alternative for approximating such structural properties, particularly the second largest eigenvalue λ_2 . This approximation is pivotal for understanding graph connectivity and is crucial in the analysis of cubic graphs, where even though all are 3-regular, their structures can be exceedingly complex.

Number of Vertices	Graphs
4	1
6	2
8	5
10	19
12	85
14	509
16	4 060
18	41 301
20	510 489
22	7 319 447
24	117 940 535
26	2 094 480 864
28	40 497 138 011
30	845 480 228 069
32	18 941 522 184 590
34	453 090 162 062 723
36	11 523 392 072 541 432
38	310 467 244 165 539 782
40	8 832 736 318 937 756 165

Table 2.4: The number of connected cubic graphs on a given number of vertices, illustrating the exponential growth in complexity with increasing vertices, see [1].

Theorem 2.1 (*First Theorem of Graph Theory*) For any graph $G = (V, E)$, the total of all vertex degrees is equal to twice the number of edges:

$$\sum_{v \in V} d(v) = 2m,$$

where $m = |E|$.

This theorem and the large number of graphs analyzed underscore why our study focuses on cubic graphs with order up to 20. This specific choice allows for a manageable yet sufficiently diverse dataset for exploring the nuances introduced

CHAPTER 3

ADVANCED INSIGHTS IN REGRESSION ANALYSIS: RELATIONSHIPS AND PREDICTIVE MODELING

Regression analysis is a statistical approach that is frequently employed in data analysis to explore and model the relationships between variables. It seeks to understand the influence of a single or multiple independent variables (often called *explanatory variables* or *predictors*) on a dependent variable (often referred to as an *outcome* or *response variables*). By applying a regression model to observed data, this method facilitates predictions, identifies patterns, and quantifies the strength and nature of relationships. Utilized across diverse fields such as economics, finance, biology, and many others, regression analysis provides valuable insights into the dynamics of com-

plex systems, making it an indispensable tool in statistical analysis.

There are several types of regression analysis, each serving different purposes based on the characteristics of the data and the nature of the relationships being explored.

Here are some common types:

- **Linear Regression:** This is the simplest type of regression, in which a linear equation represents the association between the dependent variable and the independent variable or variables.
- **Polynomial Regression:** This method is used when the relationship between variables is best described by a polynomial equation (e.g., quadratic, cubic), allowing for the modeling of nonlinear patterns.

These examples highlight the versatility of regression analysis, which includes other specialized forms adapted to particular analytical requirements and datasets. The selection of regression method relies on the research goals and the characteristics of the data.

The benefits of regression analysis in data analysis and decision-making are manifold. One major advantage is its ability to elucidate relationships between variables. It quantifies how changes in one variable correspond with changes in another, providing deep insights into the connections between them. Regression models are also pivotal for prediction; once a relationship has been established, the model can predict outcomes for new or unseen data. Additionally, by offering a quantitative understanding of relationships, regression analysis supports informed decision-making, enabling businesses and researchers to make choices grounded in robust data-driven evidence.

Overall, regression analysis stands as a powerful and widely used statistical technique that significantly contributes to informed decision-making, hypothesis testing, and deeper insights into data relationships.

3.1 Linear Regression Analysis

Linear regression is a statistical model employed to understand the association between one or more predictor variables and a response variable. For example, it can be used to examine the relationship between individuals' weights and heights. Before modeling, it is crucial to ascertain whether there exists a notable correlation between the variables involved. This correlation implies association, not causation. For instance, higher SAT scores do not necessarily cause higher college grades. A scatterplot can help visualize the relationship between variables. If no trend appears, fitting a linear regression may not be useful. The correlation coefficient, ranging from -1 to 1 , quantitatively describes the strength of the relationship.

3.1.1 Simple Linear Regression

Simple linear regression is a foundational statistical technique utilized for modeling the correlation between a single predictor variable and a response variable. This approach makes the assumption that a straight line can adequately explain the relationship between the two variables.

Model Formulation: The model for simple linear regression is written as:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i, \quad (3.1)$$

where:

- Y_i is the response variable for the i -th observation.
- X_i is the predictor variable for the i -th observation.
- β_0 and β_1 are parameters representing the intercept and slope of the regression line, respectively.
- ϵ_i represents the random error term for the i -th observation, assumed to be normally distributed with mean zero and variance σ^2 .

Estimation via Least Squares: The parameters β_0 and β_1 are typically estimated using the least squares criterion, this involves minimizing the total of squared residuals between observed responses and responses predicted by the linear model:

$$\min_{\beta_0, \beta_1} \sum_{i=1}^n (Y_i - \beta_0 - \beta_1 X_i)^2. \quad (3.2)$$

This minimization leads to the normal equations, from which the least squares estimates of the parameters are derived:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \quad (3.3)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}, \quad (3.4)$$

where $\bar{X}$ and $\bar{Y}$ are the sample means of the predictor and response variables, respectively.

Model Diagnostics and Assumptions: After fitting the model, it is crucial to check the assumptions upon which linear regression is based:

- **Linearity:** The relationship between the predictor and response should be linear. Scatter plots of residuals versus fitted values should not show patterns.
- **Independence:** Observations are independent of each other.
- **Homoscedasticity:** Constant variance of error terms regardless of the value of the predictor.
- **Normality of Residuals:** A normal distribution should be roughly followed by the residuals.

Techniques such as residual plots, Durbin-Watson statistic for detecting autocorrelation, and Shapiro-Wilk test for normality are useful tools for diagnosing potential violations of these assumptions.

Interpretation and Application: The coefficient β_1 signifies the estimated alteration in the response variable associated with a one-unit adjustment in the predictor variable, assuming all other factors are held constant. This model is widely used in various fields for predictive analysis, understanding relationships between variables, and guiding decision-making processes.

By following this structured approach, simple linear regression provides a powerful tool for data analysis, offering insights into the underlying patterns governing observed

phenomena.

3.1.2 Multiple Linear Regression

By adding more predictor variables, multiple linear regression goes beyond simple linear regression, offering a more nuanced understanding of the relationships influencing the response variable. This model is crucial when the impact of a variable is contingent upon the values of others.

Model Formulation: The mathematical model for multiple linear regression is given by:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_p X_{ip} + \epsilon_i, \quad (3.5)$$

where:

- Y_i is the response for the i -th observation.
- $X_{i1}, X_{i2}, \dots, X_{ip}$ are the predictor variables for the i -th observation.
- $\beta_0, \beta_1, \dots, \beta_p$ are the model coefficients.
- ϵ_i represents the error term for the i -th observation, assumed to be independent and normally distributed with mean zero and constant variance.

Matrix Representation: In matrix terms, the model can be compactly written as:

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}, \quad (3.6)$$

where:

- $\mathbf{Y}$ is an $n \times 1$ matrix of the response variable.
- $\mathbf{X}$ is an $n \times (p + 1)$ matrix (often called the "design matrix") that includes a column of ones for the intercept and columns for each predictor.
- $\boldsymbol{\beta}$ is a $(p + 1) \times 1$ matrix of coefficients.
- $\boldsymbol{\epsilon}$ is an $n \times 1$ matrix of errors.

The least squares estimation aims to minimize the sum of squared residuals:

$$\text{SSR} = (\mathbf{Y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{Y} - \mathbf{X}\boldsymbol{\beta}). \quad (3.7)$$

The normal equations can be solved to find the solution to this minimization problem:

$$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}. \quad (3.8)$$

From this, the estimate of $\boldsymbol{\beta}$ is given by:

$$\boldsymbol{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}, \quad (3.9)$$

assuming that $\mathbf{X}'\mathbf{X}$ is non-singular.

Diagnostic and Adequacy Checks: Following model estimation, it is essential to perform diagnostic checks to validate the model assumptions such as linearity, independence, homoscedasticity, and normality of residuals. Tools commonly used include:

- Residual plots and scatter plots to check for non-linear patterns or heteroscedasticity.
- Influence measures to identify influential data points.
- Variance inflation factor (VIF) for diagnosing multicollinearity among predictors.

Interpretation: Each coefficient β_j in a multiple regression context is interpreted to represent the change in the response variable for a one-unit change in the predictor X_j , while maintaining the constant values of all other predictors. This interpretation is critical in studies where multiple factors simultaneously impact the outcome.

This comprehensive approach in multiple linear regression allows researchers to dissect complex interdependencies among variables, providing insights that are critical in several disciplines, including social sciences, medicine, and economics.

3.1.3 Method of Least Squares

The method of least squares is a fundamental approach in statistical modeling and data analysis. Its primary objective is to ascertain the line (or hyperplane in multiple regression) that best aligns with the observed data. In order to do this, the sum of the squared differences, or residuals, between the values predicted by the model and the actual values must be minimized.

Goal of Least Squares: The least squares method aims to minimize the residu-

als' sum of squares. This objective is encapsulated in the following function:

$$Q = \sum_{i=1}^n (Y_i - (\beta_0 + \beta_1 X_i))^2, \quad (3.10)$$

where Q represents the total sum of squared residuals. By minimizing Q , the least squares method ensures that the fitted model is as close as possible to the actual data points in terms of the Euclidean distance.

Example 1: Persistence Study In a small-scale study, a researcher assigned three participants an extremely challenging task to assess their persistence. The following data were collected, including each participant's age and the number of attempts they made to finish the task before quitting:

Participant i :	1	2	3
Age X_i :	20	55	30
Attempts Y_i :	5	12	10

Table 3.1: Data from the Persistence Study

Figure 3.1: Least Squares Criterion Visualization

Least Squares Estimators: To achieve the minimization of Q , two main computational techniques are employed:

- i. **Analytical Solution:** This approach involves solving the normal equations derived from setting the partial derivatives of Q with respect to the model

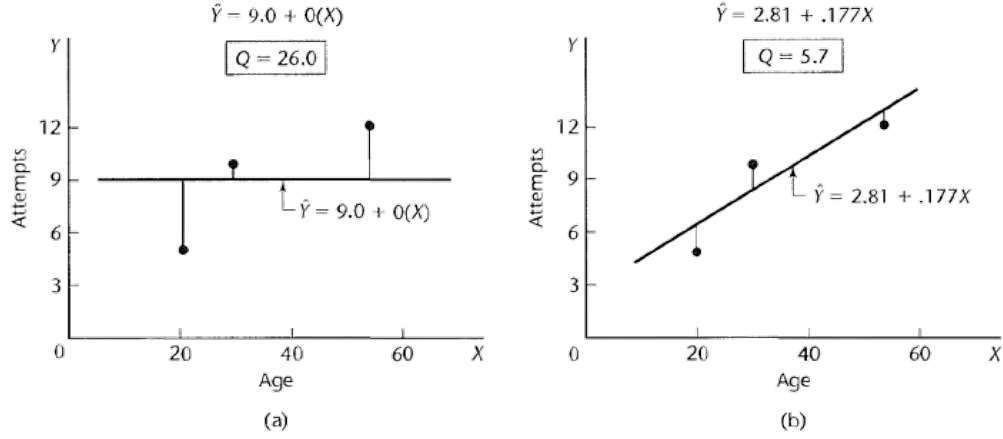


Figure 3.1: Visualization of the Least Squares Criterion Q for Regression Line Fit - Example from the Persistence Study.

parameters to zero. The normal equations are:

$$n\hat{\beta}_0 + \hat{\beta}_1 \sum X_i = \sum Y_i, \quad (3.11)$$

$$\hat{\beta}_0 \sum X_i + \hat{\beta}_1 \sum X_i^2 = \sum X_i Y_i. \quad (3.12)$$

Solving these equations yields explicit formulas for the estimates of the intercept and slope, which are:

$$\hat{\beta}_1 = \frac{\sum (X_i - \bar{X})(Y_i - \bar{Y})}{\sum (X_i - \bar{X})^2}, \quad (3.13)$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}. \quad (3.14)$$

- ii. **Numerical Optimization:** In cases where an analytical solution is not feasible (such as in non-linear models or when there are constraints on the parameters), numerical techniques like gradient descent or iterative algorithms are used to approximate the solutions. These methods iteratively adjust the parameter

estimates to find the minimum of Q .

By employing these methods, the least squares technique provides estimates of the model parameters that best explain the variability in the response variable given the predictors. The efficacy of the least squares estimators is typically assessed through model diagnostics, which examine the validity of the model assumptions and the goodness of fit.

3.1.4 Method of Maximum Likelihood

The maximum likelihood approach (MLE) is a powerful statistical tool used for parameter estimation in probabilistic models. This method focuses on selecting the values of parameter that optimize the likelihood function, indicating the likelihood of the observed data based on specific parameter values.

Goal of Maximum Likelihood Estimation: The primary objective of MLE is to identify the parameter values that render the observed data most likely according to the assumed model. This is achieved by maximizing the likelihood function, which for a regression model, is typically formulated assuming normally distributed errors.

Likelihood Function: For a simple linear regression model where the errors are assumed to be normally distributed, independently and identically distributed (i.i.d.) with mean zero and variance σ^2 , the likelihood function for the observations can be

expressed as:

$$L(\beta_0, \beta_1, \sigma^2 | x, y) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - \beta_0 - \beta_1 x_i)^2}{2\sigma^2}\right) \quad (3.15)$$

$$= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2\right). \quad (3.16)$$

The log-likelihood, which simplifies the computations, is:

$$\ln L(\beta_0, \beta_1, \sigma^2 | x, y) = -\frac{n}{2} \ln(2\pi\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - \beta_0 - \beta_1 x_i)^2.$$

Maximizing the Likelihood: To obtain the values of β_0 , β_1 , and σ^2 that maximize the likelihood function, we take partial derivatives of the log-likelihood with respect to each parameter and set them to zero. Solving these equations provides the MLEs for the parameters.

Example 2: Height and Weight Relationship Suppose a study investigates the relationship between height (in inches) and weight (in pounds) for a group of individuals. The dataset is as follows:

Participant	Height (X)	Weight (Y)
1	68	165
2	70	180
3	72	190

Table 3.2: Data from the Height and Weight Study

Using the method of maximum likelihood, we assume a linear relationship:

$$Y_i = \beta_0 + \beta_1 X_i + \epsilon_i,$$

where $\epsilon_i \sim N(0, \sigma^2)$.

After setting up and maximizing the likelihood function, we find:

$$\begin{aligned}\hat{\beta}_1 &= \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^n (X_i - \bar{X})^2}, \\ \hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}, \\ \hat{\sigma}^2 &= \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_i)^2.\end{aligned}$$

Conclusion: MLE provides a framework for estimating the parameters of a linear regression model that are most likely to have generated the observed data, under the assumption of normally distributed errors. It is particularly useful in contexts where the distribution of the errors can be assumed or is known, making the parameter estimates more interpretable and statistically efficient.

3.1.5 Method of Maximum Likelihood for Multiple Linear Regression

In the context of multiple linear regression, the method of maximum likelihood (MLE) extends naturally to accommodate multiple predictor variables. This extension allows us to estimate the parameters of a model that includes more than one explanatory variable, thereby providing a more detailed analysis of complex relationships.

Model Description: In multiple linear regression, the model is expressed as:

$$Y_i = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik} + \epsilon_i,$$

where Y_i is the dependent variable, $X_{i1}, X_{i2}, \dots, X_{ik}$ are the predictor variables, $\beta_0, \beta_1, \dots, \beta_k$ are the coefficients to be estimated, and ϵ_i are the error terms assumed to be normally distributed with mean zero and variance σ^2 .

Likelihood Function: For multiple predictors, the likelihood function for normally distributed errors is given by:

$$\begin{aligned} L(\beta_0, \beta_1, \dots, \beta_k, \sigma^2 | x, y) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}))^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - (\beta_0 + \beta_1 x_{i1} + \dots + \beta_k x_{ik}))^2\right). \end{aligned}$$

The corresponding log-likelihood function simplifies the maximization process.

Maximizing the Likelihood: The maximization of the likelihood function with respect to each parameter involves setting the partial derivatives of the log-likelihood with respect to each β_j and σ^2 to zero and solving the resulting system of equations.

Application Example: Consider a study examining the impact of both age and diet on blood pressure. The model would include age and diet as predictors, allowing us to assess their individual and combined effects on blood pressure.

This method is particularly advantageous in settings where the predictors might exhibit multicollinearity, as the MLE can provide estimates even when traditional least squares estimation might be problematic due to the inversion of nearly singular

matrices.

Conclusion: MLE in multiple linear regression offers a robust way to estimate model parameters that best describe the observed data, assuming a normal distribution of errors. It is particularly useful for understanding the relative impacts of multiple factors in a single model, providing insights into complex interdependencies.

3.1.6 Normal Equation for Linear Regression Models

The Normal Equation is a formula used for determining the parameters which minimize the cost function in both simple and multiple linear regression. It provides a direct analytic solution to fitting the linear regression model without the need for iterative optimization.

Model Representation

In linear regression, whether simple or multiple, the model can typically be expressed in a matrix form:

$$\mathbf{Y} = \mathbf{X}\beta + \epsilon,$$

where:

- $\mathbf{Y}$ is an $n \times 1$ vector representing the dependent variable.
- $\mathbf{X}$ is an $n \times (k + 1)$ matrix of k predictors (including one column of ones for the intercept if applicable).
- β is a $(k + 1) \times 1$ vector of parameters, including the intercept and slope coefficients.

- ϵ is an $n \times 1$ vector of the residuals or errors.

Derivation of the Normal Equation

To minimize the sum of squared residuals $RSS = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$, we differentiate RSS with respect to β and set the derivative to zero, yielding:

$$\mathbf{X}'\mathbf{X}\beta = \mathbf{X}'\mathbf{Y}.$$

This results in the Normal Equation:

$$\beta = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

This equation assumes that $\mathbf{X}'\mathbf{X}$ is invertible. If not, techniques like regularization might be required.

Examples in Simple and Multiple Linear Regression

Simple Linear Regression:

Consider predicting house prices based only on their size (single predictor). The dataset might include:

House Size (sq ft)	House Price (1000)
2100	400
1600	330
2400	450

For simple linear regression, $\mathbf{X}$ includes the intercept and house size:

$$\mathbf{X} = \begin{bmatrix} 1 & 2100 \\ 1 & 1600 \\ 1 & 2400 \end{bmatrix}, \quad \mathbf{Y} = \begin{bmatrix} 400 \\ 330 \\ 450 \end{bmatrix}.$$

Multiple Linear Regression:

Expanding to multiple predictors, suppose we also consider the number of bedrooms:

House Size (sq ft)	Number of Bedrooms	House Price (1000)
2100	3	400
1600	2	330
2400	4	450

The $\mathbf{X}$ matrix now includes the number of bedrooms:

$$\mathbf{X} = \begin{bmatrix} 1 & 2100 & 3 \\ 1 & 1600 & 2 \\ 1 & 2400 & 4 \end{bmatrix}.$$

Visualization

While visualizing the model fit in multiple regression is challenging due to higher dimensions, for simple regression, the fit can be depicted as:

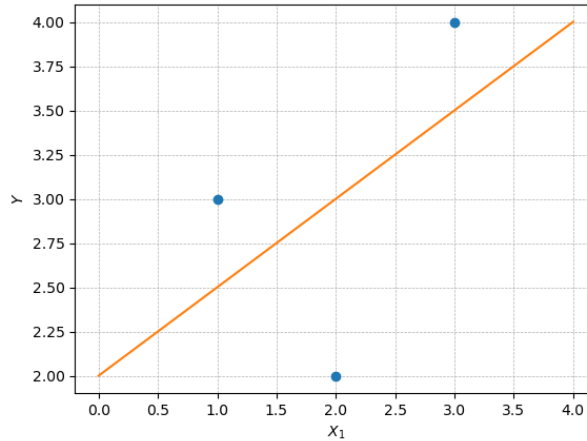


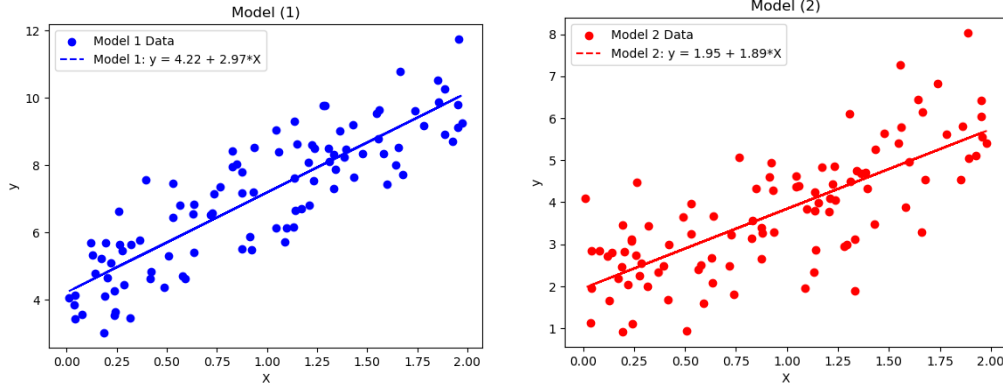
Figure 3.2: A scatter plot with original points and a regression line (orange).

This approach simplifies parameter estimation by directly solving the set of equations, particularly advantageous when dealing with smaller datasets or fewer predictors.

3.2 Coefficient of Determination

The coefficient of determination R^2 quantifies the dispersion of observations around a regression line or another statistical model. It indicates the portion of the variability in the response variable that can be attributed to changes in the explanatory variable (or variables).

R^2 for linear models is always in the range of 0 and 1. Greater spread is indicated by R^2 values closer to zero, while larger values show that the data is more densely packed around the regression line.



(a) Model (1) with $R^2 = 0.75$

(b) Model (2) with $R^2 = 0.55$

Figure 3.3: Comparison of Two Linear Regression Models

The relationship is often summarized as follows:

$$R^2 = 1 - \frac{SSR}{SST}$$

where SSR represents the sum of squared residuals and SST is the total sum of squares. It should be mentioned that R^2 might be the statistic in statistics that people misunderstand the most. It is crucial to be mindful of what it says and does not say. R^2 quantifies the extent to which the variability in y can be accounted for by variability in x ; as per its definition, it tends to be lower for datasets with higher $\text{var}(y_i)$. Models with R^2 near 1 are not necessarily good. For instance: the linear model shown below has $R^2 = 0.95$ despite fitting the data rather poorly.

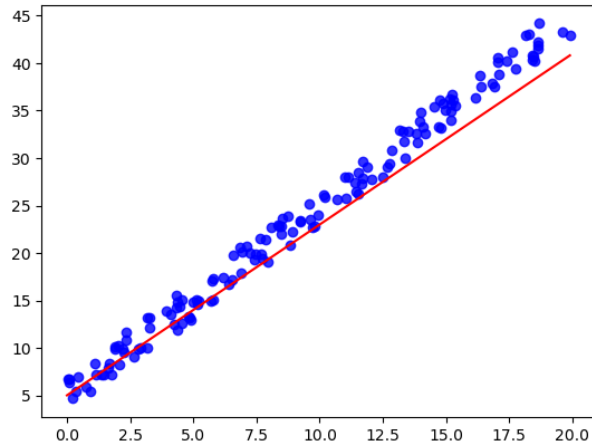


Figure 3.4: A model with high R - squared

Similarly, models with low R^2 are not necessarily bad. The below linear model fits the data fairly well, yet $R^2 = 0.66$ because the data is naturally noisy.

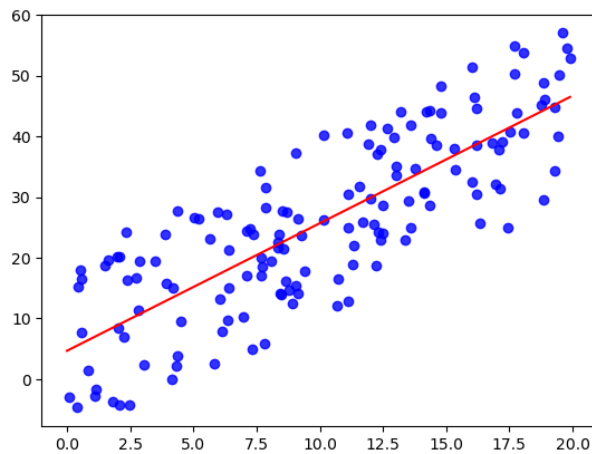


Figure 3.5: A model with low R - squared

Thus, the coefficient of determination measures variability about a regression line. It does not directly tell if the model is useful or even reasonable.

CHAPTER 4

REFINED MODELING AND DATA ANALYSIS TECHNIQUES

This chapter details the data extraction processes and the development of predictive models, specifically focusing on modeling the relationship between the Wiener index of cubic graphs and their second-largest eigenvalues.

4.1 Data Extraction

To estimate λ_2 , the second-largest eigenvalue of the adjacency matrix, we utilize cubic graphs' Wiener indices. Data from *The House of Graphs* [13] provides these indices for cubic graphs in (.g6) format, which includes adjacency lists. We use *Mathematica* to convert these lists into matrices necessary for our computations.

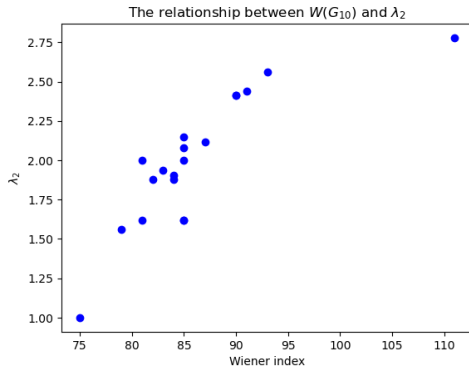
4.2 Development of the Regression Model

Our goal is to establish a regression model that predicts λ_2 based on the Wiener index.

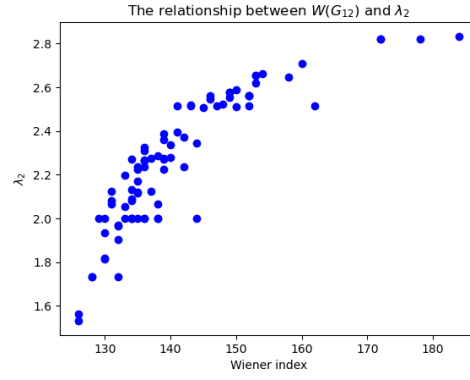
Initial data exploration focused on understanding the relationship between λ_2 and the Wiener index, adjusting for the number of vertices in the graphs.

4.2.1 Data Visualization

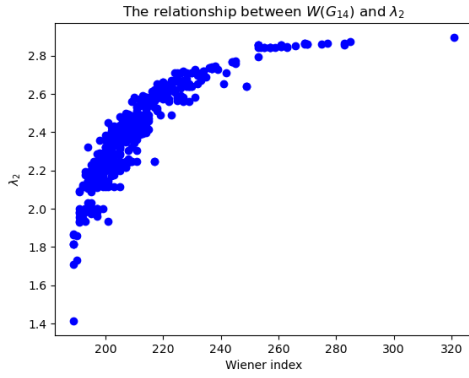
The relationship between the Wiener indices for cubic graphs of different sizes and λ_2 is visualized in Figure 4.1.



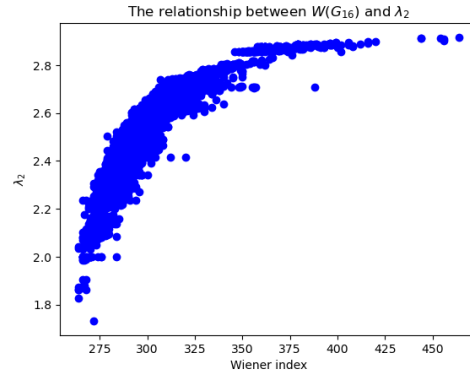
(a) G_{10} : a cubic graph with $n = 10$.



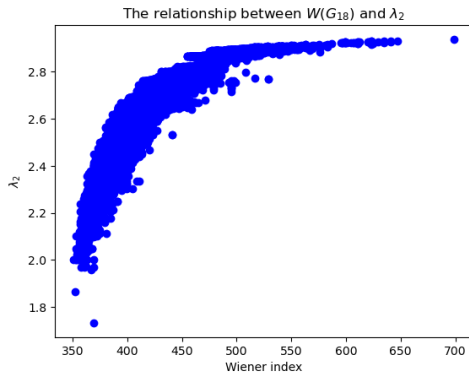
(b) G_{12} : a cubic graph with $n = 12$.



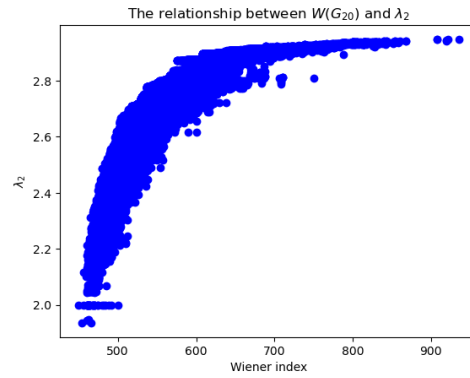
(c) G_{14} : a cubic graph with $n = 14$.



(d) G_{16} : a cubic graph with $n = 16$.



(e) G_{18} : a cubic graph with $n = 18$.



(f) G_{20} : a cubic graph with $n = 20$.

Figure 4.1: Scatter plots illustrating the relationship between the Wiener indices of cubic graphs and λ_2 across different orders.

4.2.2 Exploration of Regression Models for Cubic Graphs of Orders 10 to 20

In our analysis, we explored a range of regression models to accurately predict the second-largest eigenvalue λ_2 for cubic graphs. We investigated linear to higher-degree polynomial models, including quadratic, cubic, and quartic models, as well as square root and logarithmic functions to assess their performance. The performance of each model was evaluated based on their R-squared values, which measure the proportion of variance in λ_2 that can be explained by the models.

Below, the R-squared values for different models for cubic graphs of orders 10 to 20 are provided in a series of tables:

Function	Terms	R-Squared
Linear	x	0.6874
Quadratic	$x + x^2$	0.8459
Square-root	$\sqrt{x}$	0.7103
	$x + \sqrt{x}$	0.8450
	$x + x^2 + \sqrt{x}$	0.8454
	$x^2 + \sqrt{x}$	0.8456
Logarithmic	$\log x$	0.7319
	$x + \log x$	0.8446
	$x + x^2 + \log x$	0.8454
	$x^2 + \log x$	0.8452
Quartic	$x + x^2 + x^3 + x^4$	0.8705

Table 4.1: Comparing the R-Squared for different models of G_{10}

Function	Terms	R-Squared
Linear	x	0.7430
Quadratic	$x + x^2$	0.8547
Square-root	$\sqrt{x}$	0.7591
	$x + \sqrt{x}$	0.8577
	$x + x^2 + \sqrt{x}$	0.8643
	$x^2 + \sqrt{x}$	0.8559
Logarithmic	$\log x$	0.7740
	$x + \log x$	0.8586
	$x + x^2 + \log x$	0.8644
	$x^2 + \log x$	0.8570
Quartic	$x + x^2 + x^3 + x^4$	0.8648

Table 4.2: Comparing the R-Squared for different models of G_{12}

Function	Terms	R-Squared
Linear	x	0.7352
Quadratic	$x + x^2$	0.8711
Square-root	$\sqrt{x}$	0.7565
	$x + \sqrt{x}$	0.8793
	$x + x^2 + \sqrt{x}$	0.8933
	$x^2 + \sqrt{x}$	0.8743
Logarithmic	$\log x$	0.7762
	$x + \log x$	0.8816
	$x + x^2 + \log x$	0.8935
	$x^2 + \log x$	0.8773
Quartic	$x + x^2 + x^3 + x^4$	0.8951

Table 4.3: Comparing the R-Squared for different models of G_{14}

Function	Terms	R-Squared
Linear	x	0.7470
Quadratic	$x + x^2$	0.8802
Square-root	$\sqrt{x}$	0.7684
	$x + \sqrt{x}$	0.8880
	$x + x^2 + \sqrt{x}$	0.9009
	$x^2 + \sqrt{x}$	0.8833
Logarithmic	$\log x$	0.7879
	$x + \log x$	0.8901
	$x + x^2 + \log x$	0.9011
	$x^2 + \log x$	0.8862
Quartic	$x + x^2 + x^3 + x^4$	0.9021

Table 4.4: Comparing the R-Squared for different models of G_{16}

Function	Terms	R-Squared
Linear	x	0.7508
Quadratic	$x + x^2$	0.8774
Square-root	$\sqrt{x}$	0.7718
	$x + \sqrt{x}$	0.8851
	$x + x^2 + \sqrt{x}$	0.8978
	$x^2 + \sqrt{x}$	0.8805
Logarithmic	$\log x$	0.7908
	$x + \log x$	0.8872
	$x + x^2 + \log x$	0.8980
	$x^2 + \log x$	0.8833
Quartic	$x + x^2 + x^3 + x^4$	0.8991

Table 4.5: Comparing the R-Squared for different models of G_{18}

Function	Terms	R-Squared
Linear	x	0.7573
Quadratic	$x + x^2$	0.8721
Square-root	$\sqrt{x}$	0.7767
	$x + \sqrt{x}$	0.8790
	$x + x^2 + \sqrt{x}$	0.8905
	$x^2 + \sqrt{x}$	0.8749
Logarithmic	$\log x$	0.7942
	$x + \log x$	0.8809
	$x + x^2 + \log x$	0.8907
	$x^2 + \log x$	0.8774
Quartic	$x + x^2 + x^3 + x^4$	0.8917

Table 4.6: Comparing the R-Squared for different models of G_{20}

4.2.3 Model Validation for Cubic Graphs of Order 10

This subsection is dedicated to validating the regression model developed to predict the second-largest eigenvalue, λ_2 , specifically for cubic graphs of order 10. The model employed is a quartic polynomial, chosen for its robust performance in initial tests.

Model Formulation

The proposed model predicts λ_2 based on the Wiener index (x) using the following quartic polynomial equation:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$

Here, y denotes the predicted λ_2 , and x is the Wiener index of the graph.

Parameter Estimation

The model coefficients ($\beta_0, \beta_1, \beta_2, \beta_3$, and β_4) are computed using the least squares method, which involves solving this matrix equation:

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{Y}$$

Matrix Configuration

- **Design Matrix $\mathbf{X}$:** Each row corresponds to $[1, x_i, x_i^2, x_i^3, x_i^4]$, representing polynomial terms for the Wiener index of the i -th graph:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix}$$

- **Response Vector $\mathbf{Y}$:** Comprises the observed λ_2 values:

$$\mathbf{Y} = \begin{bmatrix} \lambda_{2,1} \\ \lambda_{2,2} \\ \vdots \\ \lambda_{2,n} \end{bmatrix}$$

Computation

Coefficients are derived by solving the matrix equation using inversion techniques, provided $\mathbf{X}^T \mathbf{X}$ is invertible.

Visualization and Model Assessment

Effective visualization and assessment of the model are crucial for validating the predictive accuracy and identifying any potential biases.

Figure 4.2 illustrates the predicted λ_2 values against the Wiener Index, highlighting the model's accuracy and fit.

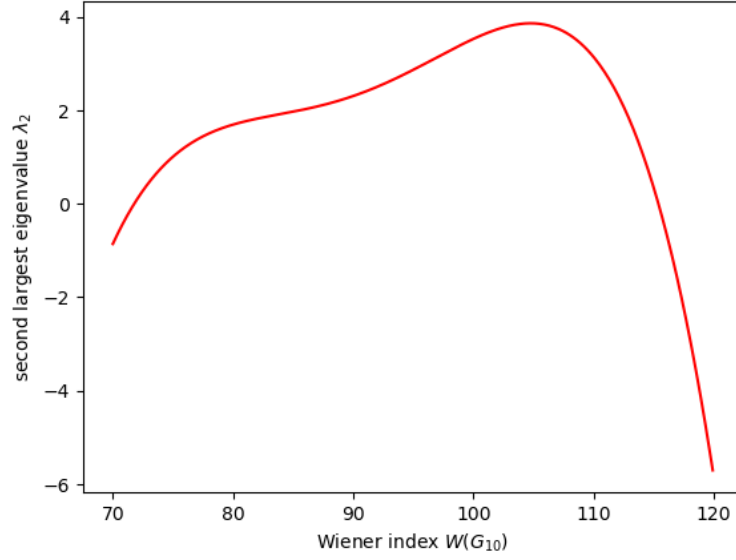
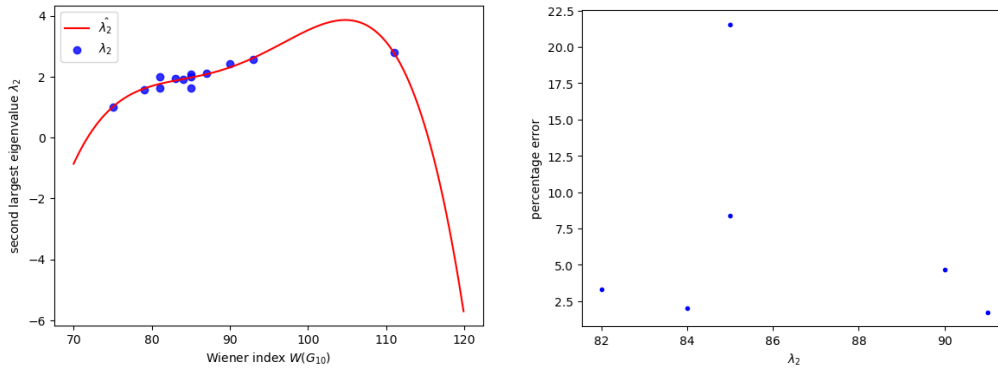


Figure 4.2: Correlation between the Wiener index and the estimated second-largest eigenvalue λ_2 for cubic graphs of order 10.

Figure 4.3 provides an in-depth look at both the actual versus predicted values and the distribution of residuals, which helps in assessing the model's predictive consistency and the validity of its assumptions.



(a) Actual versus predicted λ_2 values. (b) Percentage error in predictions.

Figure 4.3: Analysis of model performance based on training and testing data for cubic graphs of order 10.

These graphical tools play a pivotal role in the model validation process, providing

clear insights into the effectiveness of the regression approach and pinpointing areas where the model might be improved.

4.2.4 Model Validation for Cubic Graphs of Orders 12, 14, 16, 18, 20

This subsection validates the regression models designed to predict the second-largest eigenvalue, λ_2 , for cubic graphs of various orders. Each order employs a quartic polynomial model, which was selected based on its effectiveness during preliminary testing phases.

Model Formulations and Parameter Estimations: For each order, the relationship between the Wiener index x and λ_2 is modeled using a quartic polynomial equation:

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \beta_3 x^3 + \beta_4 x^4$$

where y represents the predicted λ_2 and x denotes the Wiener index of the graph.

The coefficients ($\beta_0, \beta_1, \beta_2, \beta_3$, and β_4) for each model are determined through the least squares method by solving the matrix equation:

$$\mathbf{X}^T \mathbf{X} \boldsymbol{\beta} = \mathbf{X}^T \mathbf{Y}$$

Model Specifics for Each Order:

Order 12:

$$y = -225.671959 + 5.53055626x - 0.0504629098x^2 + 0.000205241134x^3 - 3.13371223 \times 10^{-7}x^4$$

Order 14:

$$y = -79.9948343 + 1.2139703x - 0.00671345486x^2 + 1.65669571 \times 10^{-5}x^3 - 1.53531084 \times 10^{-8}x^4$$

Order 16:

$$y = -61.4700197 + 0.657693325x - 0.00253678917x^2 + 4.36617890 \times 10^{-6}x^3 - 2.82248730 \times 10^{-9}x^4$$

Order 18:

$$y = -49.6094781 + 0.389569173x - 0.00108821905x^2 + 1.35297247 \times 10^{-6}x^3 - 6.30046632 \times 10^{-10}x^4$$

Order 20:

$$y = -46.8012325 + 0.284224016x - 0.000611879477x^2 + 5.86185380 \times 10^{-7}x^3 - 2.10257164 \times 10^{-10}x^4$$

Matrix Setup: For each model, the design matrix $\mathbf{X}$ and the response vector $\mathbf{Y}$

are constructed as follows:

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 & x_1^2 & x_1^3 & x_1^4 \\ 1 & x_2 & x_2^2 & x_2^3 & x_2^4 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ 1 & x_n & x_n^2 & x_n^3 & x_n^4 \end{bmatrix}$$

$$\mathbf{Y} = \begin{bmatrix} \lambda_{2,1} \\ \lambda_{2,2} \\ \vdots \\ \lambda_{2,n} \end{bmatrix}$$

Visualization and Model Assessment:

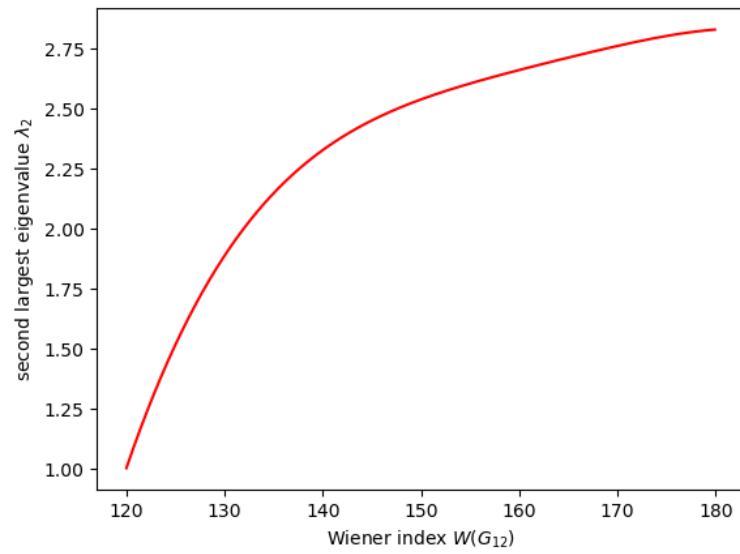
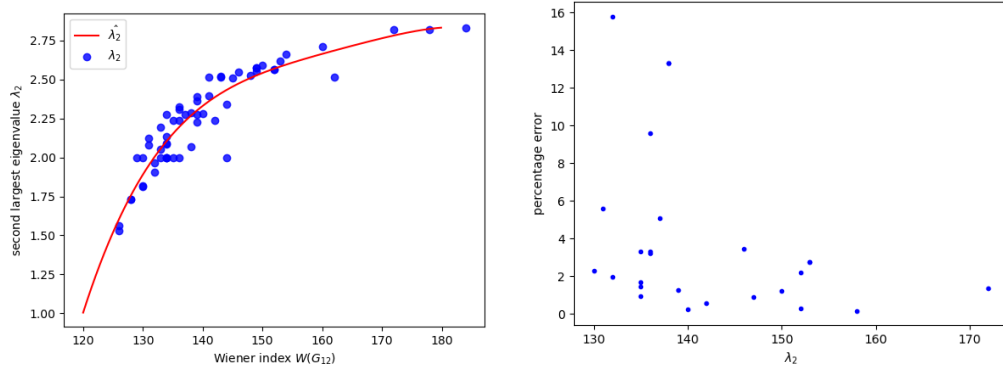
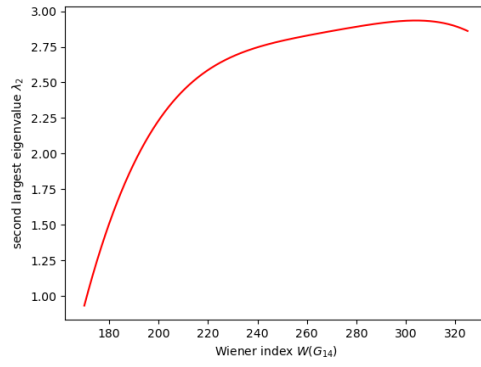


Figure 4.4: Exploration of the connection between Wiener's index and λ_2 of cubic graphs of order 12

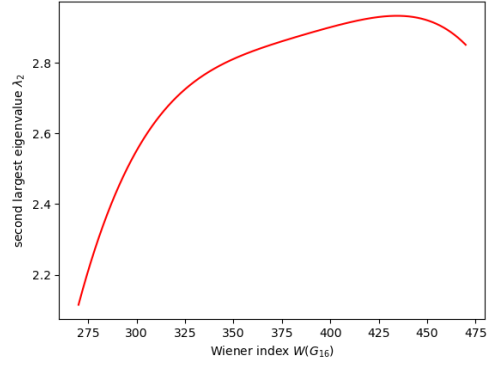


(a) The actual and the fitted values. (b) The percentage error of the testing data.

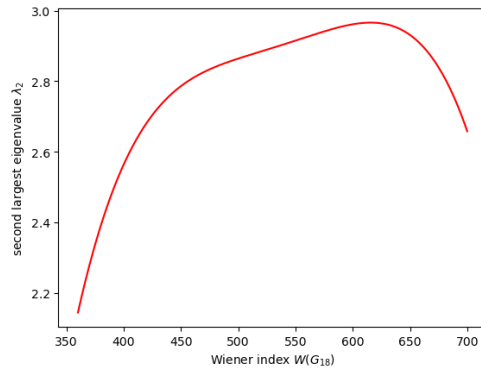
Figure 4.5: Indications derived from the training and testing datasets for order 12.



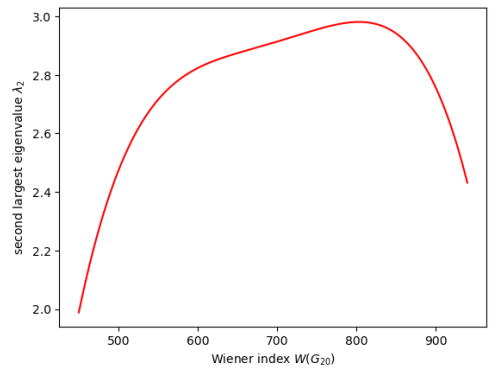
(a) Order 14



(b) Order 16

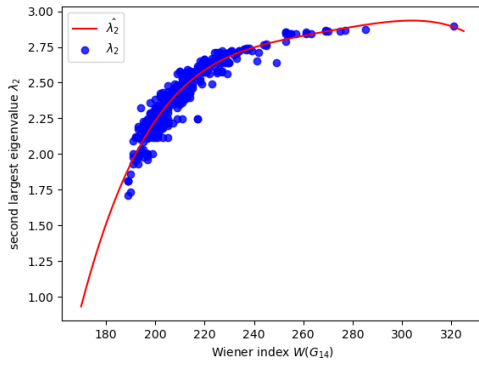


(c) Order 18

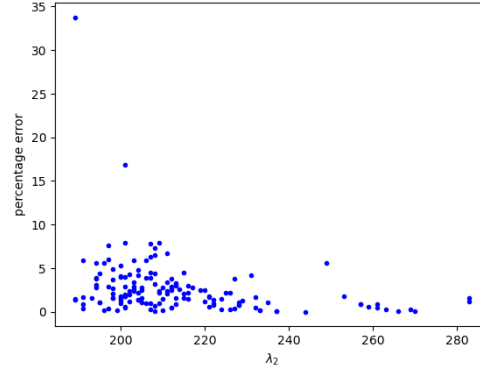


(d) Order 20

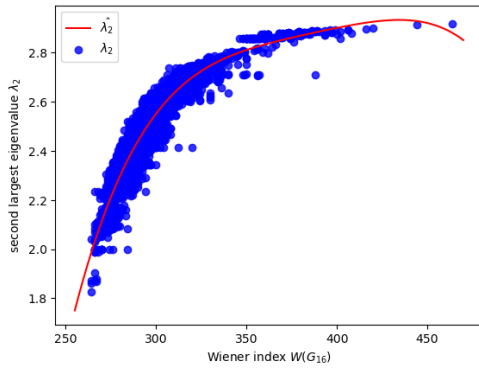
Figure 4.6: Models from the parameters obtained from the corresponding dataset for orders 14, 16, 18, and 20.



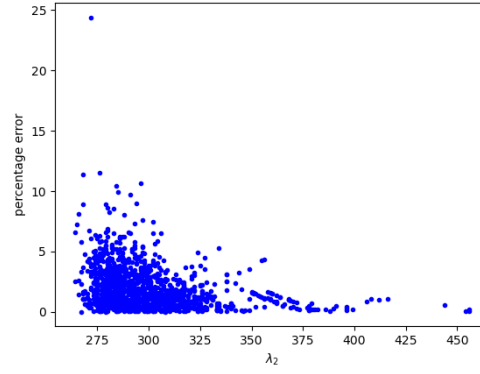
(a) Order 14 - Actual vs. Predicted λ_2



(b) Order 14 - Percentage Error

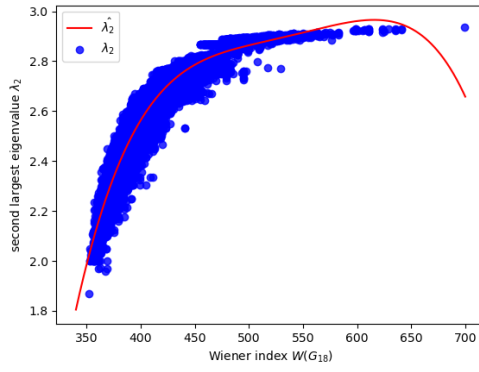


(c) Order 16 - Actual vs. Predicted λ_2

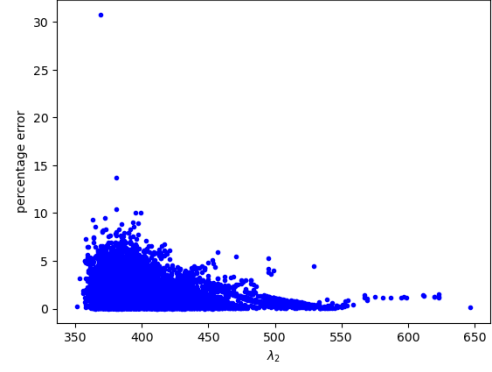


(d) Order 16 - Percentage Error

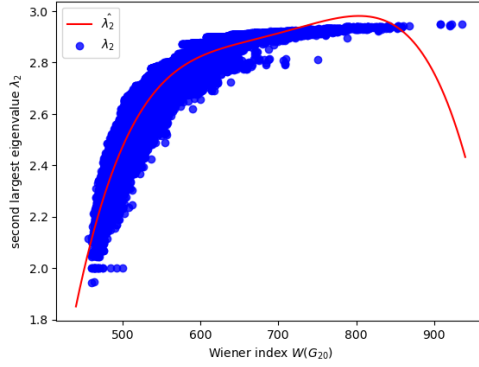
Figure 4.7: Predictions of the training and testing data for orders 14 and 16.



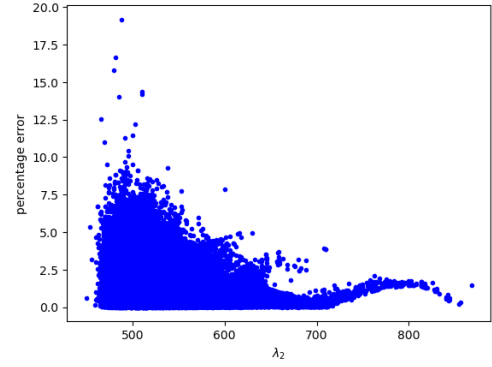
(a) Order 18 - Actual vs. Predicted λ_2



(b) Order 18 - Percentage Error



(c) Order 20 - Actual vs. Predicted λ_2



(d) Order 20 - Percentage Error

Figure 4.8: Predictions of the training and testing data for orders 18 and 20.

These visual representations offer critical insights into both the efficacy and the potential biases or errors inherent in the predictive framework, highlighting the model's accuracy and areas for improvement.

4.2.5 Practical Application

This subsection demonstrates the practical application of our regression models by analyzing cubic graphs of various orders to validate the model's predictive capability across a diverse dataset. Each graph order was tested to assess the accuracy of the

model in predicting the second-largest eigenvalue, λ_2 .

Here is a detailed breakdown of the performance metrics for each graph order:

- **Order 10:** We examined 19 non-isomorphic cubic graphs. The model's predictions showed that the percentage error for the majority was below 5%, indicating high precision in smaller graph sizes.
- **Order 12:** Out of 85 non-isomorphic graphs analyzed, approximately 94% of the dataset exhibited errors of less than or equal to 5%, confirming the model's robustness.
- **Order 14:** For the 509 non-isomorphic cubic graphs, around 88% of the testing data points fell within the acceptable error range of less than 5%, demonstrating consistent performance.
- **Order 16:** With 4060 graphs evaluated, a significant 94% of the predictions maintained an error below 5%, showcasing the model's capability to handle larger datasets effectively.
- **Order 18:** Involving 41301 non-isomorphic graphs, this order showed exceptional accuracy, with 97% of the predictions not exceeding the 5% error threshold.
- **Order 20:** The largest set analyzed included 510489 non-isomorphic graphs, where about 99% of the model's predictions stayed within the 5% error margin, highlighting excellent predictive power even for extensive graph orders.

These results validate the model's utility in practical scenarios, where accurate predictions of λ_2 are crucial across varying graph sizes and complexities. The detailed analysis confirms the model's reliability and adaptability, reinforcing its application in real-world graph theory challenges.

4.2.6 Average of the Wiener Index

This section explores the significance of the average Wiener index in predicting the second-largest eigenvalue, λ_2 , for cubic graphs across various orders. It provides insights into how data centered around the average index affects the predictive accuracy of the model.

Methodological Overview

The study initially focused on graphs of order 12 due to limited data for order 10. We computed the average Wiener index, μ_n , for each class and categorized the data accordingly:

- Graphs with $W(G_n) \geq \mu_n$
- Graphs with $W(G_n) > \mu_n$
- Graphs with $W(G_n) \leq \mu_n$
- Graphs with $W(G_n) < \mu_n$

Application of the Quartic Model

For cubic graphs of order 12 with $W(G_{12}) \geq 141$ (33 out of 85 data points), we applied the following quartic model:

$$y = -927.064218 + 227.917x - 20.9251138x^2 + 0.852401153x^3 - 0.000129939899x^4$$

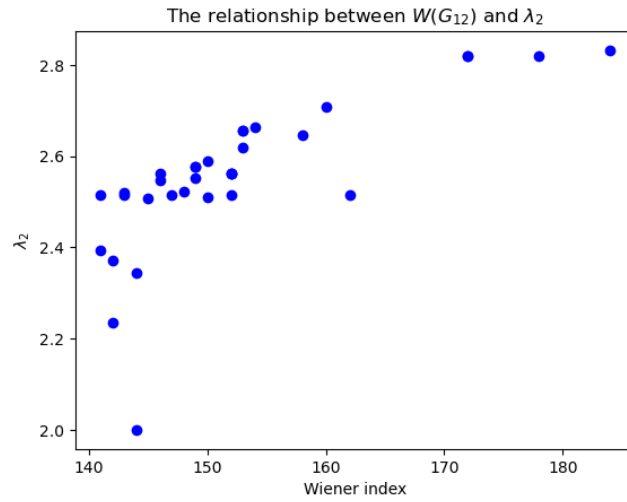
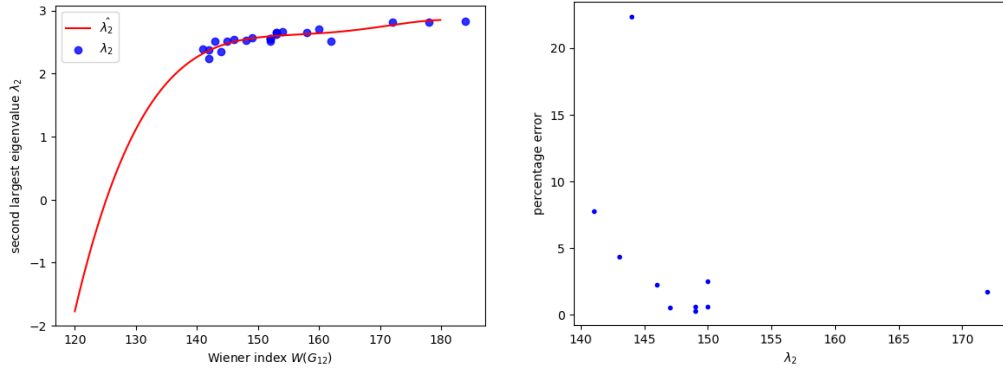


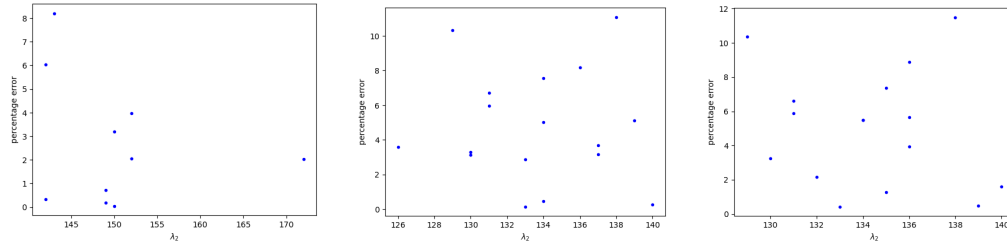
Figure 4.9: Exploration of the connection between Wiener's index and λ_2 for cubic graphs of order 12 with Wiener indices $\geq$ the average.



(a) Actual vs. predicted values

(b) Percentage error distribution

Figure 4.10: Model performance and error distribution for cubic graphs of order 12. These plots are critical for evaluating the accuracy of the regression model and for understanding the variability of the model's predictions compared to actual data.



(a) $W(G_{12}) > \mu_{12}$

(b) $W(G_{12}) \leq \mu_{12}$

(c) $W(G_{12}) < \mu_{12}$

Figure 4.11: The percentage error of the testing data for different cases. This figure breaks down the error percentage based on whether the Wiener index is greater than, less than, or equal to the average, providing deeper insight into model performance across these subgroups.

Detailed Analysis of Additional Orders

Expanding the analysis to include higher orders provided further validation of the model's robustness:

Order 14:

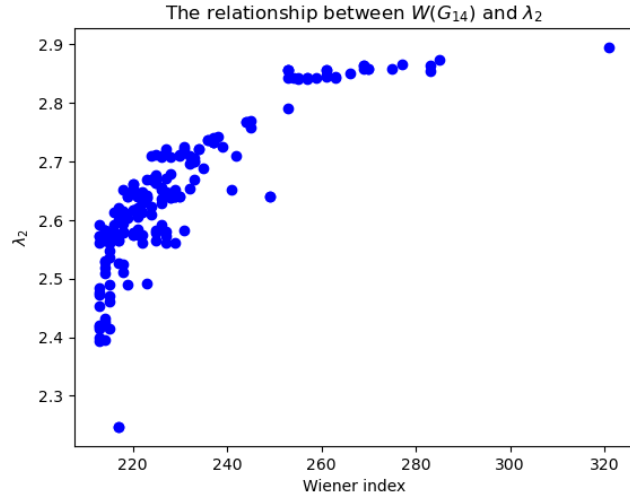


Figure 4.12: Exploration of the connection between Wiener's index and λ_2 of cubic graphs of order 14, for the dataset with $W(G_{14}) \geq 213$.

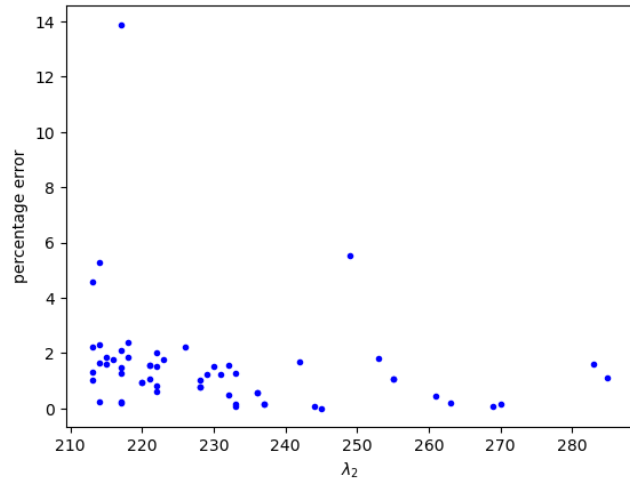


Figure 4.13: The percentage error of the testing data for cubic graphs of order 14, highlighting precision in model predictions.

Order 16:

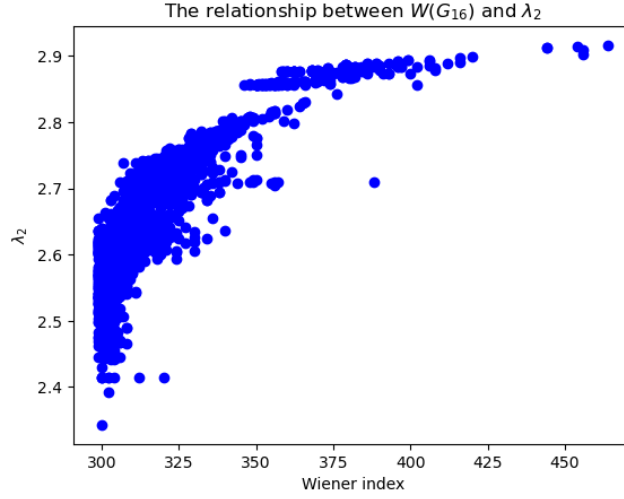
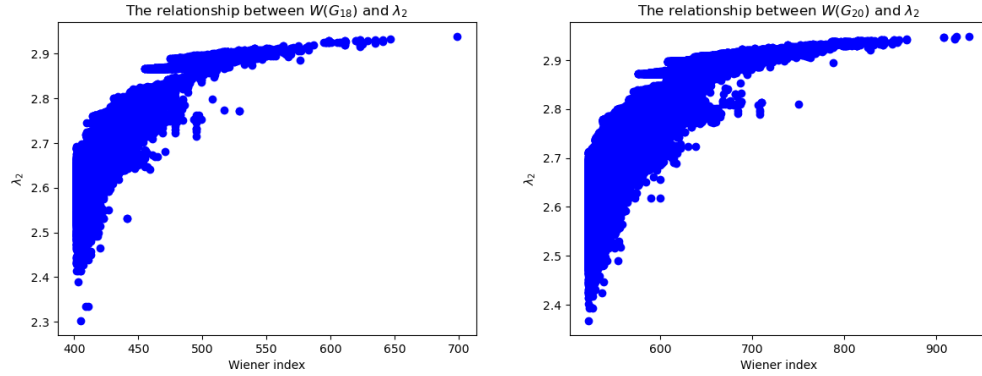


Figure 4.14: The relation between the data points where $W(G_{16}) \geq 299$ and λ_2 , illustrating the model's accuracy for high Wiener indices.

General Exploration:



(a) $W(G_{18}) \geq \mu_{18}$: Detailed exploration
 (b) $W(G_{20}) \geq \mu_{20}$: Detailed exploration

Figure 4.15: Exploration of the connection between Wiener's index and λ_2 across different cubic graph orders, confirming consistency and reliability of the predictive model.

Comprehensive Summary and Model Efficacy Across Orders

Enhanced Accuracy by Considering Averages The analysis underscores that segmenting the data based on the average Wiener indices markedly enhances the model's accuracy. Specifically, for cubic graphs where the Wiener index meets or exceeds the average, the models demonstrate exceptionally low percentage errors. For instance, approximately 94% of the testing data for the class of 18 vertices and about 95% for the class of 20 vertices exhibited a percentage error of less than or equal to 3%. This precise categorization based on average Wiener indices not only confirms the robustness of our models but also highlights the critical role of considering average values in the predictive analysis of graph properties.

Model Performance Across Different Graph Orders Across all examined graph orders, the majority of the testing data consistently showed percentage errors below 3%, evidencing the models' high precision:

- **Order 12:** Around 70% of the testing data points had percentage errors of less than or equal to 3%.
- **Order 14:** Approximately 93% of the testing data achieved errors within this range.
- **Order 16:** Similarly, 93% of the data points exhibited errors of less than or equal to 3%.
- **Order 18 and 20:** Showed percentage errors of less than or equal to 3% in about 94% and 95% of the testing data, respectively.

These results demonstrate the effectiveness and reliability of the quartic polynomial models in accurately predicting the second largest eigenvalues, highlighting their suitability for broad applications in graph theory and network analysis.

Conclusion

This comprehensive study has validated the effectiveness of using the Wiener index to predict the second-largest eigenvalue in cubic graphs across multiple orders. The consistent accuracy and reliability across varying conditions affirm the utility of our modeling approach for theoretical and practical applications in graph theory and related fields.

CHAPTER 5

MODEL GENERALIZATION FOR CUBIC GRAPHS

This chapter details the development and validation of a comprehensive model for predicting the second-largest eigenvalue of cubic graphs, utilizing the Wiener index as a predictor.

5.1 Development of a Generalized Regression Model

We initiated our analysis by plotting data for cubic graphs with vertex orders $n = 10, 12, 14, 16, 18$, and 20 . This exploratory analysis is fundamental to understanding the underlying patterns and adjusting our modeling strategy accordingly.

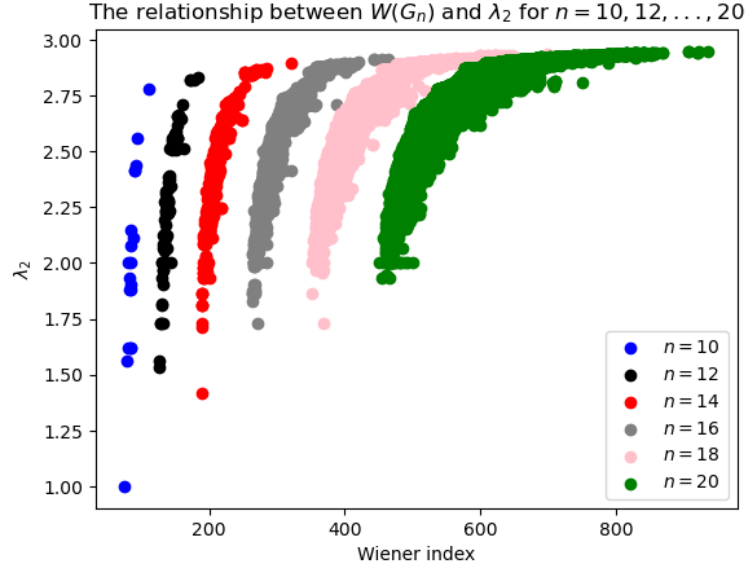


Figure 5.1: Scatter plot exploring the relationship between the Wiener index and λ_2 across cubic graphs of various orders

As demonstrated in Figure 5.1, the data exhibits consistent patterns with a noticeable shift towards higher values as the vertex count n increases. Consequently, we solidified our regression model for cubic graphs with a vertex order of $n = 20$, as follows:

$$\begin{aligned}
 y_{20} = & (-4.68012325 \times 10) + (2.84224016 \times 10^{-1}) x - (6.11879477 \times 10^{-4}) x^2 \\
 & + (5.86185380 \times 10^{-7}) x^3 - (2.10257164 \times 10^{-10}) x^4.
 \end{aligned} \tag{5.1}$$

To utilize this model, we adjust the Wiener index of each graph under examination, denoted as G . This adjustment involves aligning G 's Wiener index to reflect the average differences within its class. Specifically, we incorporate the deviation between μ_{20} , the average for the $n = 20$ class, and μ_G , the average for the graph's class.

We now apply this refined model to predict the second-largest eigenvalue for graphs

of class $n = 14$, adjusting each graph's Wiener index as $W(G) + |\mu_{20} - \mu_{14}|$. Below, we present the percentage errors resulting from these predictions.

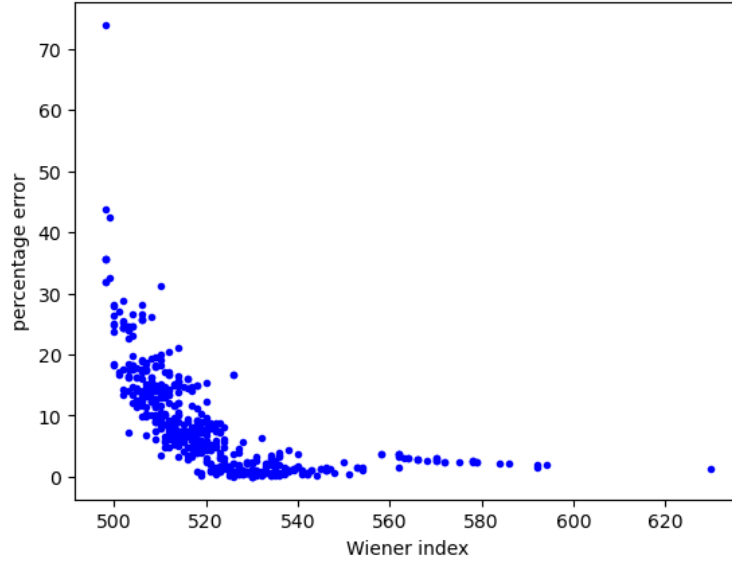
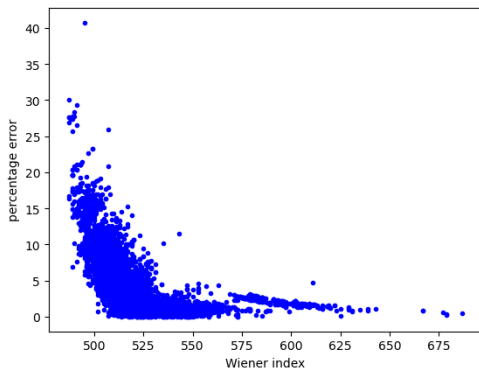
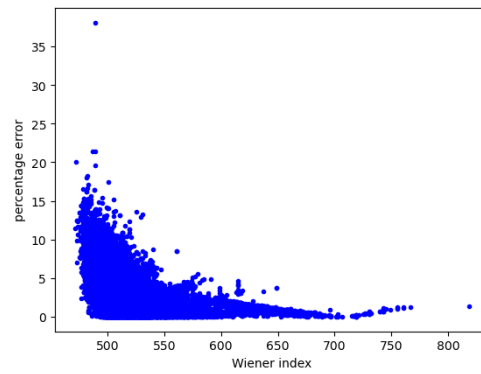


Figure 5.2: Analysis of prediction errors for the second-largest eigenvalue in graphs of order 14

Further, we extend this model to predict the second-largest eigenvalues for graphs with $n = 16$ and 18.



(a) Analysis of prediction errors for the
class of 16



(b) Analysis of prediction errors for the
class of 18

Figure 5.3: Evaluation of the generalized regression model's performance for orders 16 and 18

The error analyses in Figure 5.3 illustrate that the prediction errors for λ_2 generally remain below 7.5

5.2 Enhanced Predictive Modeling Using Wiener Index Averages

In our ongoing efforts to refine the predictions of the second-largest eigenvalues, we employed a strategy similar to that discussed in Chapter 4, which focuses on incorporating the average Wiener index in the approximation process. We started by analyzing data from all graphs with vertex orders ranging from $n = 10$ to 20, specifically including graphs where the Wiener indices exceed the average for their respective classes. The analysis is visualized in the following graph:

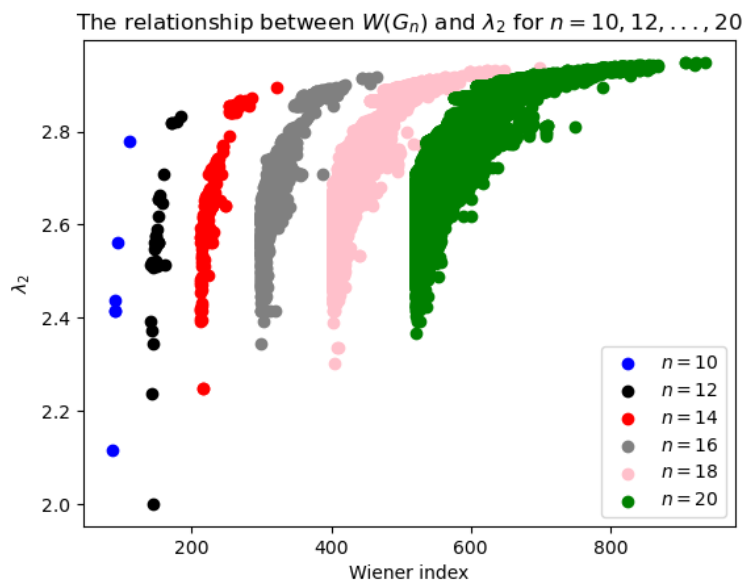


Figure 5.4: Analysis of the relationship between Wiener's index and λ_2 for cubic graphs G of various orders n , where $W(G) \geq \mu_n$

As depicted in Figure 5.4, the scatter plot consistently reflects the trends observed

in the complete dataset. Based on this consistency, we adjusted the Wiener index of each graph G using our previously established model y_{20} . This approach allows us to refine our predictions by considering the average Wiener index of each class, thereby enhancing the accuracy of the model. The established model is as follows:

$$y_{20} = (-1.84963163 \times 10) + (1.09676973 \times 10^{-1}) x - (2.12221172 \times 10^{-4}) x^2 + (1.83724407 \times 10^{-7}) x^3 - (5.99491751 \times 10^{-11}) x^4. \quad (5.2)$$

We initiated the prediction of the second-largest eigenvalue for graphs in the category with $n = 14$ utilizing the refined model y_{20} . Within this application, each x value is adjusted to $W(G) + |\mu_{20} - \mu_{14}|$ for every graph G within this category.

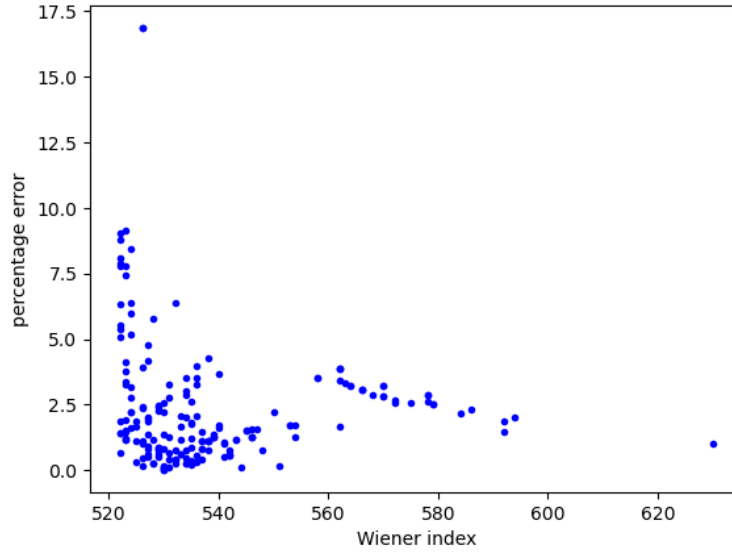
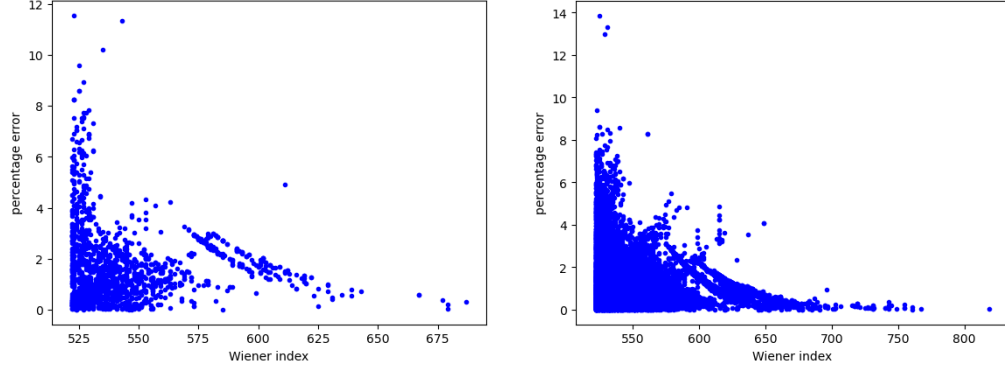


Figure 5.5: Evaluation of percentage error in predicting the second-largest eigenvalue for graphs of order 14

Figure 5.5 demonstrates that the percentage error in predicting λ_2 for graphs of order 14 generally remains below 4%.

We will extend this predictive approach to graphs with $n = 16$ and 18 using the same refined model y_{20} .



(a) Analysis of percentage error for the class of 16 (b) Analysis of percentage error for the class of 18

Figure 5.6: Assessment of the percentage error in estimating the second-largest eigenvalue for graphs of orders 16 and 18

As shown in Figure 5.6, the scatter plot on the left indicates that the percentage error for estimating λ_2 for graphs of order 16 is less than or equal to 4% for 93% of the data points. Similarly, the scatter plot on the right illustrates the percentage error when estimating λ_2 for graphs of order 18 which remains at or below 4% for approximately 97% of the dataset.

5.2.1 Exploring Wiener Indices Below the Average

To improve our approach in estimating the second-largest eigenvalue, particularly for cases where the corresponding Wiener indices are below the average μ_n , we employed Equation 5.2. This equation is specifically utilized for graphs of order n where $W(G) < \mu_n$, aiming to provide more accurate approximations.

Initially, we analyzed graphs with 14 vertices to assess the accuracy of our model:

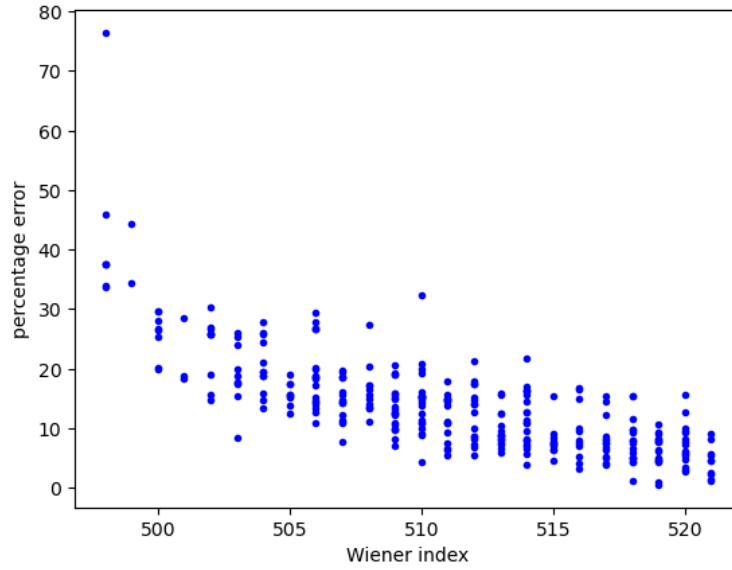
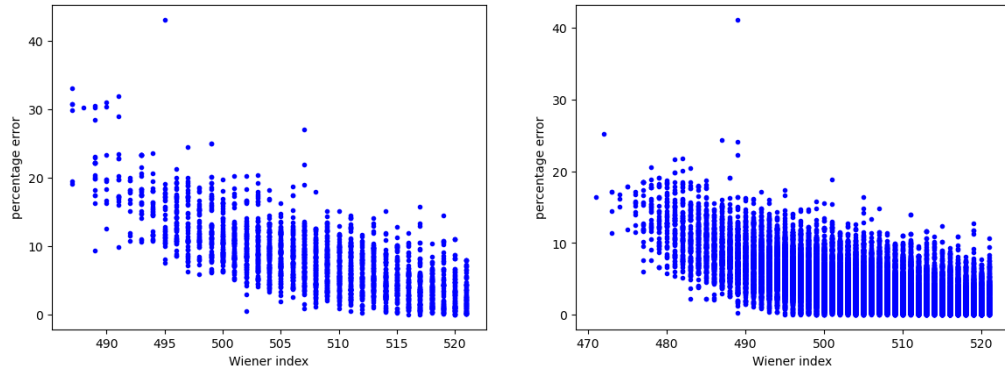


Figure 5.7: Analysis of percentage error in estimating the second-largest eigenvalue for graphs of order 14 with Wiener indices below the average

Figure 5.7 shows that for these graphs, the percentage error mostly remains below 20%, indicating that the model performs good for this subset.

To further validate our model, we extended the analysis to include graphs with 16 and 18 vertices, using the same equation to estimate λ_2 . The results are depicted in the following figures:



(a) Percentage error for graphs of order 16 (b) Percentage error for graphs of order 18

Figure 5.8: Evaluation of percentage errors in estimating the second-largest eigenvalues for graphs with 16 and 18 vertices

As shown in Figure 5.8, the percentage error for these orders remains significantly below 10% for most data points, supporting the effectiveness of the equation across a range of graph orders.

Following these encouraging results, we applied the model to estimate λ_2 for graphs with 20 vertices, specifically targeting those with Wiener indices below the average. This iteration of the model incorporated additional parameters to enhance its predictive accuracy:

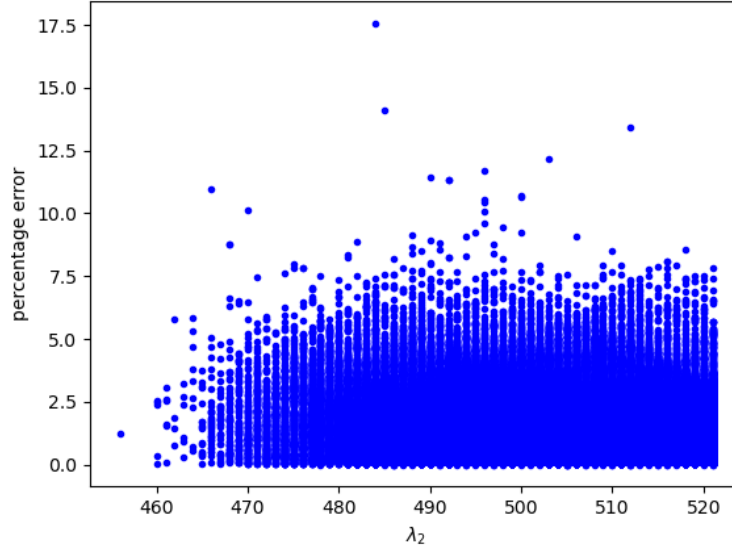


Figure 5.9: The percentage error for estimating the second-largest eigenvalue for graphs with 20 vertices using a model focused on graphs with Wiener indices below the average

As depicted in Figure 5.9, this more tailored approach resulted in a significant reduction in percentage error, often below 5%, confirming the model's capability to accurately predict eigenvalues for graphs under these specific conditions.

This comprehensive evaluation demonstrates the variability in model performance across different graph orders and emphasizes the necessity of adapting the methodological approach to improve prediction accuracy in graph analysis.

CHAPTER 6

CONCLUSION AND RECOMMENDATIONS

In this chapter, we present our conclusions on approximating the second largest eigenvalue using the Wiener index and provide recommendations for future research directions.

6.1 Conclusion

Our study has yielded the following conclusions regarding the approximation of the second largest eigenvalue:

1. We have successfully utilized the Wiener index to approximate the second largest eigenvalue for cubic graphs with orders up to 20, achieving a percentage error of less than or equal to 5%.
2. By focusing on graphs with a Wiener index greater than or equal to the average

Wiener index of each specific class, we have obtained improved approximations with a percentage error of less than or equal to 3%.

3. We successfully developed a generalized model for cubic graphs of up to order 20, achieving a percentage error of 7.5% or less.
4. We effectively derived a generalized model that incorporates the average, enhancing performance for cubic graphs of order 20 or lower, and achieving a percentage error of less than or equal to 4%.

6.2 Recommendations for Future Work

Based on our findings, we propose the following recommendations for future research:

1. Explore the applicability of the Wiener index approach to graph classes beyond cubic graphs. Investigate its performance with different graph structures and properties to assess its versatility and robustness.
2. Study the sensitivity of the Wiener index-based approach to various parameters or graph properties. Identify the factors that have the most significant influence on the accuracy of the eigenvalue approximation.
3. Generalize the approach by incorporating indices that consider both the graph's degree and the distances between its vertices. Extend the applicability beyond cubic graphs of orders up to 20 to encompass a broader range of cubic graphs.
4. Utilize indices that integrate the graph's degree and vertex distances to generalize the approach not only for cubic graphs but also for all k -regular graphs

with $k \geq 4$.

5. Investigate the integration of machine learning techniques into the Wiener index approach to enhance predictive accuracy, especially when dealing with complex graph structures or large datasets. Explore the potential of leveraging machine learning algorithms to improve the efficiency and scalability of the eigenvalue approximation method.

These recommendations aim to expand the scope and applicability of the Wiener index approach, paving the way for further advancements in eigenvalue approximation techniques and their practical utility in various domains.

REFERENCES

- [1] R. W. Robinson and N. C. Wormald, “Numbers of cubic graphs,” *Journal of Graph Theory*, vol. 7, no. 4, pp. 463–467, 1983.
- [2] A. Dobrynin, R. Entringer, and I. Gutman, “Wiener index of trees: Theory and applications,” *Acta Applicandae Mathematicae*, vol. 66, pp. 211–249, 2001.
- [3] I. Gutman, “Selected properties of the schultz molecular topological index,” *Journal of Chemical Information and Computer Sciences*, vol. 34, no. 5, pp. 1087–1089, 1994.
- [4] S. Akhter, “Two degree distance based topological indices of chemical trees,” *IEEE Access*, vol. 7, pp. 95 653–95 658, 2019.
- [5] F. Alsaadi, S. Ahtsham, S. Bokhary, A. Shah, U. Ali, J. Cao, M. Alassafi, M. Rehman, and J. Ul Rahman, “On knowledge discovery and representations of molecular structures using topological indices,” *Journal of Artificial Intelligence and Soft Computing Research*, vol. 11, p. 21, 2021.
- [6] M. Adnan, S. Bokhary, and M. Imran, “On wiener polarity index and wiener index of certain triangular networks,” *Journal of Chemistry*, vol. 2021, pp. 1–20,

2021.

- [7] M. Imran, S. Akhter, and Z. Iqbal, “Edge mostar index of chemical structures and nanostructures using graph operations,” *International Journal of Quantum Chemistry*, vol. 120, 2020.
- [8] H. Wiener, “Structural determination of paraffin boiling points,” *Journal of the American Chemical Society*, vol. 69, no. 1, pp. 17–20, 1947.
- [9] H. Hosoya, “Topological index. a newly proposed quantity characterizing the topological nature of structural isomers of saturated hydrocarbons,” *Bulletin of The Chemical Society of Japan - BULL CHEM SOC JPN*, vol. 44, pp. 2332–2339, 1971.
- [10] M. Randic, “Characterization of molecular branching,” *Journal of the American Chemical Society*, vol. 97, no. 23, pp. 6609–6615, 1975.
- [11] A. Balaban, “Highly discriminating distance-based topological index,” *Chemical Physics Letters*, vol. 89, no. 5, pp. 399–404, 1982.
- [12] A. T. Balaban, “Topological indices based on topological distances in molecular graphs,” *Pure and Applied Chemistry*, vol. 55, no. 2, pp. 199–206, 1983.
- [13] S. D. K. Coolsaet and J. Goedgebeur. (2023) House of graphs 2.0: A database of interesting graphs and more. [Online]. Available: <https://houseofgraphs.org/>

VITAE

- Name: Shahad Abdulaziz Aldossary.
- Nationality: Saudi.
- Date of Birth: Feb 3, 1999.
- Email: *ShahadAbdulaziz35@gmail.com*
- Permenant Address: 8354, 6A Street, Tihama neighborhood, 4685-34255
Dhahran, Kingdom of Saudi Arabia.

Education:

- M.Sc. in Mathematics from King Fahd University of Petroleum and Minerals (KFUPM), Dhahran, Saudi Arabia, May 2024. Thesis title: "Wiener Index Estimations of Second-Largest Eigenvalues Across Cubic Graph Orders".
- B.Sc. in Applied Mathematics from Imam Abdulrahman Bin Faisal University (IAU), Dammam, Saudi Arabia, May 2021.

Qualification:

- **King Fahd University of Petroleum and Minerals | 2022 - Present:**
Graduate Assistant.
- **Saad National Schools (International Department) | 2021 - 2022:**
Mathematics Teacher.