https://dspace.org/

Connected domination number, average distance and their application in drug design

Flowers, James David John

2025

University of the Witwatersrand, Johannesburg

https://hdl.handle.net/10539/50254

Flowers, James David John. (2025). Connected domination number, average distance and their application in drug design. [Master's dissertation, University of the Witwatersrand, Johannesburg]. WIReDSpace. https://hdl.handle.net/10539/50254

Downloaded from a DSpace repository. https://dspace.org/

Connected domination number, average distance and their application in drug design

Master's Dissertation

A dissertation submitted to the School of Mathematics under the Faculty of Science at the University of the Witwatersrand, Johannesburg, in fulfillment of the requirements for the degree of M.Sc. Mathematics.

Student: James David John Flowers
Student Number: 1887043
Supervisor: Professor Simon Mukwembi
Co-Supervisors: Professor Elizabeth Jonck & Doctor Ronald John Maartens
Year: 2025

Declaration of Authorship

I, James David John Flowers, student number 1887043, declare that the dissertation titled "Connected Domination Number, Average Distance and their Application in Drug Design"and the work presented in it is my own. I confirm that:

- The work was done wholly or mainly while in candidatum for a master's degree at this University.
- Where any part of this dissertation has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the dissertation is based on work done by myself or jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed

James David John Flowers

On the 24th of October 2025 at the University of the Witwatersrand.

Abstract

Let a graph with n vertices be G. A leaf of G is a vertex of degree 1 in G. The leaf number of G, denoted $L(G)$, is the number of leaves in a spanning tree of G that contains the maximum number of leaves possible. The Wiener index of G, denoted $W(G)$, is defined to be the sum of the distances between all unordered pairs of vertices in G. When the Wiener index of G, of order n, is divided by $\binom{n}{2}$, the average distance, denoted $\mu(G)$, is obtained. We are concerned with bounds on $W(G)$ and $L(G)$. Also, we attempt to model the biological activity of some compounds using these two Graph Theory concepts. The output of the model is analysed using statistical methods and the suitability of the parameters is determined.

Contents

1 Introduction **6**
- 1.1 Introduction . . . 6
- 1.2 Graph Theory . . . 6
- 1.3 Graph Domination and Leaf Number . . . 13
- 1.4 Distance Measures . . . 15
- 1.5 Vertex Unions on Graphs . . . 17
- 1.6 Organic Chemistry and Graph Theory . . . 17
- 1.7 Overview of Dissertation . . . 19
- 1.8 Conclusion . . . 20

2 Literature Review **21**
- 2.1 Introduction . . . 21
- 2.2 Leaf Number . . . 21
- 2.3 Wiener Index . . . 24
- 2.4 Graph Theory and Chemistry . . . 26
- 2.5 Conclusion . . . 27

3 Connected Domination Number of Graphs **28**
- 3.1 Introduction . . . 28
- 3.2 Lower Bounds . . . 28
- 3.3 Upper Bound . . . 37
- 3.4 Bounds on the Connected Domination Number . . . 38
- 3.5 Conclusion . . . 39

4 Average Distance in Graphs **40**
- 4.1 Introduction . . . 40
- 4.2 Vertex Unions in Graphs . . . 40
- 4.3 Bounds on the Wiener Index for Unicyclic Graphs . . . 41
  - 4.3.1 The Lower Bound . . . 41
  - 4.3.2 The Upper Bound . . . 51
- 4.4 Bounds on the Wiener Index for Tricyclic Graphs . . . 67
  - 4.4.1 The Lower Bound . . . 67
  - 4.4.2 The Upper Bound . . . 69
- 4.5 Bounds on the Wiener Index for t-Cyclic Graphs . . . 72
- 4.6 Bounds on Average Distance . . . 74
  - 4.6.1 Unicyclic Graphs . . . 74

4.6.2 Tricyclic Graphs 75
4.7 Conclusion 76

5 Average Distance and Leaf Number as Predictors for Biological Properties 78
5.1 Introduction 78
5.2 Hypothesis 79
5.3 Methodology 81
5.4 Results 83
5.5 Conclusion 87

6 Conclusion 88

Appendix A 91

List of Figures

1.1 Graph G. 7
1.2 A walk of G. 8
1.3 A path of G. 8
1.4 Graph P_4 and graph C_4. 9
1.5 Graph F. 9
1.6 The graph $K_{1,6}$. 10
1.7 A spanning tree. 10
1.8 A graph B. 11
1.9 Two induced subgraphs. 12
1.10 A bicyclic graph with 3 cycles. 13
1.11 A tricyclic graph with 5 cycles. 13
1.12 A graph G. 14
1.13 A spanning tree T of G. 14
1.14 Graph G. 15
1.15 Graph G_1 and graph G_2. 17
1.16 A vertex union of G_1 and G_2. 17
1.17 The chemical compound Quercetin. 18
1.18 The molecular graph for Quercetin. 18

2.1 Graph $PK_{6,8}$. 26

3.1 Initial subtree of Theorem 3.1, Case 3. 32
3.2 Operation $O1$ and Operation $O2$. 32

4.1 Process 4.1 graphs. 42
4.2 Process 4.2 graphs. 44
4.3 Graph $G^0(n,t)$. 48
4.4 A graph in $\mathcal{G}_{6,2}$. 49
4.5 Process 4.3 graphs. 52
4.6 Graphs of Case (i). 54
4.7 Process 4.4 graphs. 57
4.8 Process 4.5 initial graphs. 59
4.9 Process 4.5 final graphs. 59
4.10 Process 4.6 graphs. 62
4.11 Graph $S_{n,\lfloor t\rfloor,\lceil t\rceil}$. 65
4.12 $S_{7,0,1}$. 66
4.13 A graph in ${}^4\mathrm{G}_{n,3}$. 68

4.14 The set $\mathcal{M}$. 68
4.15 $J(6)$. 69
4.16 Graph P_6. 73
4.17 Graph $P_{6,1}$. 73
4.18 Graph $P_{6,2}$. 73
4.19 Graph $P_{6,3}$. 73
4.20 Graph $P_{6,4}$. 73

5.1 The molecular form of each compound. 80
5.2 The molecular graph of Quercetin. 82
5.3 Scatterplot of IC_{50} value against the transformed average distance. 85
5.4 Scatterplot of IC_{50} value against the leaf number. 86

List of Tables

5.1 Table of compounds and their IC_{50} values. 81
5.2 Table of compounds with calculated leaf number and average distance. . . . 83
5.3 Relationship of IC_{50} value and leaf number. 84
5.4 Relationship of IC_{50} value, average distance and transformed average distance. 84
5.5 Models for predicting the IC_{50} value. 85
5.6 Refined transformed average distance model. 87

Chapter 1

Introduction

1.1 Introduction

This dissertation presents an examination of two graph concepts and analyses their suitability as model parameters for predicting the cytotoxicity activity of certain flavonoids, defined on page 79. The concepts are the connected domination number, defined on page 13, and the average distance. In particular, the lower bound and the upper bound of the connected domination number, when the minimum degree and order are given, is investigated through the leaf number, defined on page 14. The lower bound and the upper bound to the average distance of unicyclic graphs and tricyclic graphs are investigated.

Ahmed, Moawad, Owis, Abouzid and Ahmed in [1], provide the data set that is used later on in an attempt to develop a model that predicts compounds' cytotoxicity towards human hepatocarcinoma (HepG2).

In this chapter basic concepts from Graph Theory and Chemistry are presented. This chapter is concluded by a presentation of the breakdown of the dissertation.

1.2 Graph Theory

A *graph* G is an ordered pair (V, E) where the elements of V are the *vertices* of G and E is a possibly empty set of 2-element subsets of V whose elements are called *edges*. It

is standard to use the notation $V(G)$ and $E(G)$. The graph is depicted visually by representing elements of $V(G)$ by points and elements of $E(G)$ by lines connecting these points determined by which two elements of $V(G)$ are used to form the edge. The *order* of G is the number of vertices in G given by $|V(G)| = n$, and the *size* of G is the number of edges in G given by $|E(G)| = m$. The edge $\{u, v\}$ is also denoted by uv.

If vertices, say u and v, of a graph G are joined by the edge uv in G, then u and v are *adjacent* or are *neighbours* in G. The *open neighborhood* of a vertex v, denoted $N(v)$, is the set of all neighbours of v. The *closed neighbourhood* of a vertex v, denoted $N[v]$, is $N[v] = N(v) \cup \{v\}$. Let $A \subseteq V(G)$. The *closed* and *open* neighbourhoods of A are defined as $N[A] = \cup_{v \in A} N[v]$ and $N(A) = \cup_{v \in A} N(v)$, respectively.

Example 1.2.1. *Vertex set, edge set and neighbourhoods.*

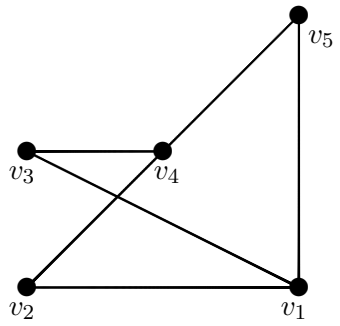


Figure 1.1: Graph G.

The graph G presented in Figure 1.1 has vertex set $V(G) = \{v_1, v_2, v_3, v_4, v_5\}$ and edge set $E(G) = \{v_1v_2, v_1v_3, v_1v_5, v_2v_4, v_3v_4, v_4v_5\}$. Thus, G has order $n = 5$ and size $m = 6$. The vertex v_1 is adjacent to all vertices except v_4, and $N(v_5) = \{v_1, v_4\}$. For $A = \{v_1, v_4\}$, note that $N[A] = \{v_1, v_2, v_3, v_4, v_5\} = V(G)$. □

For a vertex v in G the *degree* of v is $deg_G(v) = |N(v)|$, and δ denotes the *minimum degree* among all vertices in G. If $deg_G(v) = \delta$ for all vertices in G, then we call G a *δ-regular* graph. A vertex v having $deg_G(v) = 1$ is called a $leaf$, also known as an *end vertex*. When the graph is understood $deg(v)$ is used.

Example 1.2.2. *Degrees and minimum degree.*

Considering Figure 1.1,

- $deg(v_1) = deg(v_4) = 3$, and
- $deg(v_2) = deg(v_3) = deg(v_5) = 2$.

No vertex is a leaf and it is clear that $\delta = 2$. The graph is not regular. □

For vertices, say u, $v \in V(G)$, not necessarily distinct, a $u-v$ *walk* W in G is a sequence of vertices and edges in G which starts at u and ends at v such that the consecutive vertices in W are adjacent. A walk whereby each vertex appears only once in the sequence is called a *path*. Join the first and last vertices of a path to obtain a *cycle*. The *length* of a walk W is $|W| - 1$. Similarly, the length of a path and cycle is defined. Note that a cycle $(v_1, v_2, \cdots, v_n, v_1)$ is an ordered (n+1)-tuple. This dissertation, however, considers two cycles the same, when the ordering of the vertices present in the cycles are shifted. Cycle (v_2, v_1, v_3, v_4) and cycle (v_3, v_4, v_2, v_1) in Figure 1.2 below are the same.

Example 1.2.3. *Walks, paths, cycles and length.*

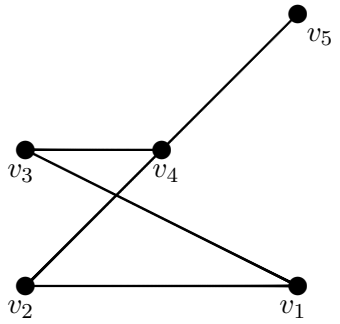


Figure 1.2: A walk of G.

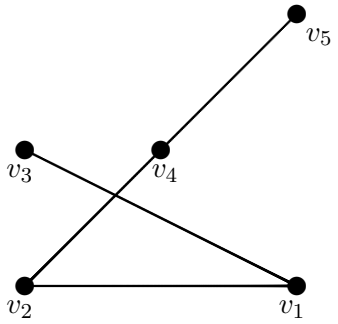


Figure 1.3: A path of G.

Figure 1.2 depicts a walk, W, of G given by $W = (v_5, v_4, v_2, v_1, v_3, v_4)$. The length of W is 5. Removing the final vertex, v_4, from the sequence of the walk results in a path of length 4. This path is depicted in Figure 1.3. Due to the vertex v_4 appearing twice in W, a cycle is contained in W. Further, this cycle is of length 4. □

Two common sets of graphs that are mentioned throughout the dissertation is a path graph and a cycle graph. The *path graph* on n vertices, denoted P_n, is defined to have $V(P_n) = \{v_1, v_2, \cdots, v_n\}$ and $E(P_n) = \{v_i v_{i+1} \,|\, 1 \leq i \leq n-1\}$. The *cycle graph* on n vertices, denoted C_n, is defined similarly to P_n, but the edge $v_n v_1$ is added. The graph P_4 and the graph C_4 are shown in Figure 1.4.

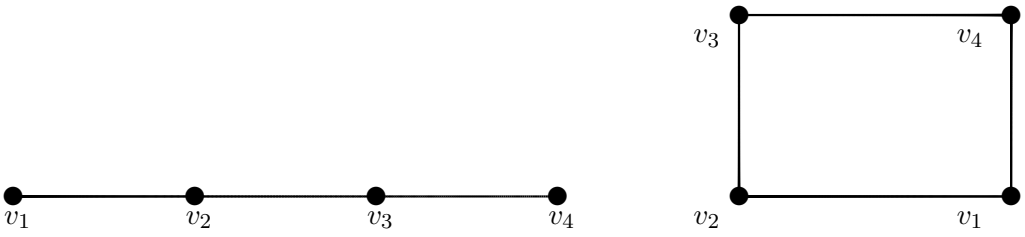


Figure 1.4: Graph P_4 and graph C_4.

If G contains a $u - v$ path then the vertices u and v are *connected* in G. If all pairs of vertices of G are connected then G is *connected*. A graph that is not connected is *disconnected*. The graph obtained from G by removing a vertex u, and all the edges joined to u, is denoted by $G - \{u\}$. If the graph G is connected and $G - \{v\}$ is disconnected, then v is called a *cut-vertex*. A connected graph containing no cut-vertices is called *non-seperable*.

A connected graph T containing no cycles is called a *tree*. For any tree T, $|E(T)| = |V(T)| - 1$. Let $T_1, \cdots, T_k$ be pairwise disconnected trees. Then, the graph F having $|V(F)| = \cup_{i=1}^{k} V(T_i)$ and $|E(F)| = \cup_{i=1}^{k} E(T_i)$ is called a *forest.*

Example 1.2.4. *Forest and tree.*

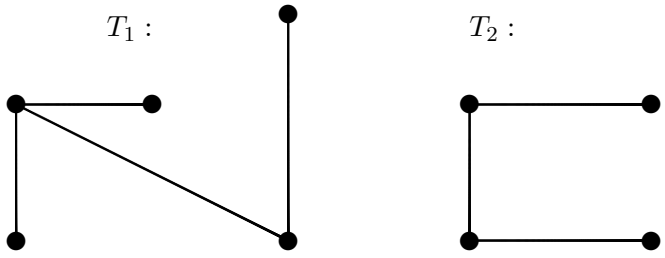


Figure 1.5: Graph F.

Forest F in Figure 1.5 consists of the 2 disconnected trees, T_1 and T_2. The trees T_1 and T_2 contains 3 and 2 leaves, respectively. □

Let a tree T have $V(T) = \{v, u_1, \cdots, u_{n-1}\}$. If $deg_T(v) = n - 1$ and $deg_T(u_i) = 1$ for $1 \leq i \leq n - 1$, then T is called the *star graph* and is denoted $K_{1,n-1}$. Shown in Figure 1.6 is the star graph on 7 vertices.

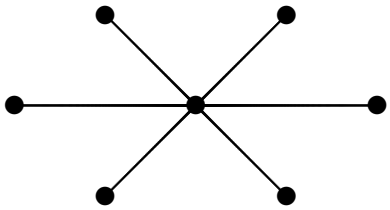


Figure 1.6: The graph $K_{1,6}$.

If H is a graph such that $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$, then H is called a *subgraph* of G. A subgraph H of G having $V(H) = V(G)$ is said to be a *spanning* subgraph. If a spanning subgraph H of a graph G is a tree, then H is called a *spanning tree* of G. Figure 1.7 depicts a spanning tree of graph G in Figure 1.1.

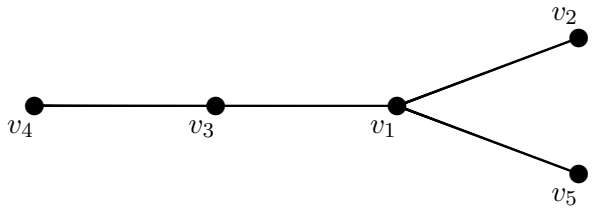


Figure 1.7: A spanning tree.

For a nonempty subset S of $V(G)$, $\langle S \rangle$ is used to denote the subgraph of G *induced* by S where $V(\langle S \rangle) = S$ and two vertices u, $v \in V(\langle S \rangle)$ are adjacent in $\langle S \rangle$ if and only if $uv \in E(G)$. A subgraph H of a graph G is called an *induced* subgraph of G if there is a nonempty set S of $V(G)$ such that $H = \langle S \rangle$. If G contains vertices $v_0, v_1, \cdots, v_n$ such that:

- the path $(v_0, v_1, \cdots, v_n)$ is a subgraph of G,
- $deg_G(v_0) \geq 3$,
- $deg_G(v_1) = 2 = deg_G(v_2) = 2 = \cdots = deg_G(v_{n-1}) = 2$, and
- $deg_G(v_n) = 1$,

then the subgraph induced by $\{v_0, v_1, \cdots, v_n\}$ is called a *pendant path*. A *block* of a graph G is a maximal nonseperable subgraph of G. The symbol, $\subseteq$, used to denote subsets is extened for graph inclusion. For example, $G_1 \subseteq G_2$ means G_1 is a subset of G_2.

Example 1.2.5. *Graph connectivity, cut-vertex and block.*

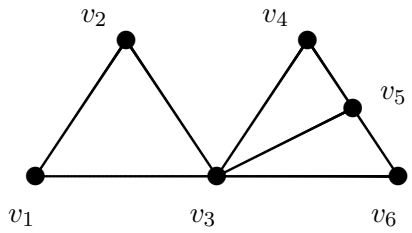


Figure 1.8: A graph B.

Figure 1.8 a depicts graph B having $V(B) = \{v_1, v_2, \cdots, v_6\}$ and $E(B) = \{v_1v_2, v_1v_3, v_2v_3, v_3v_4, v_3v_5, v_3v_6, v_4v_5, v_5v_6\}$. Clearly every pair of vertices are joined by some path, and so B is connected. Removing vertex v_3 disconnects B making v_3 a cut-vertex of B. Further, v_3 is the only cut-vertex of B. The subgraph $\langle\{v_3, v_4, v_5, v_6\}\rangle$ is nonseperable and maximal, and is therefore a block. □

Two graphs G and H are said to be *isomorphic*, denoted $G \cong H$, if there exists a bijective function $\phi : V(G) \to V(H)$ such that two vertices, say u and v, are adjacent in G if and only if $\phi(u)$ and $\phi(v)$ are adjacent in H. If no such function ϕ exists, then G and H are said to be *non-isomorphic*.

Example 1.2.6. *Subgraphs and isomorphism.*

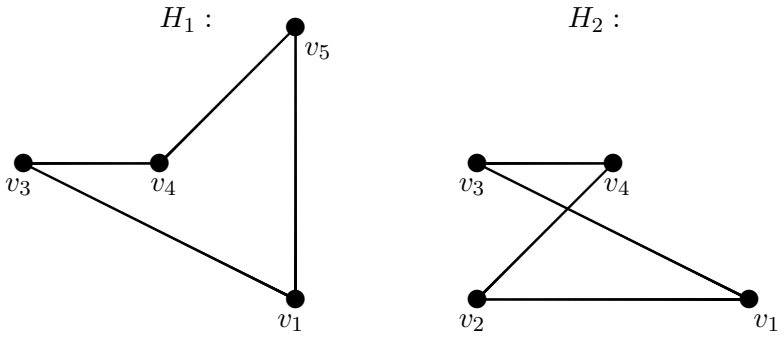


Figure 1.9: Two induced subgraphs.

Figure 1.9 shows two subgraphs of a graph G shown in Figure 1.1. Here, $H_1 = \langle\{v_1, v_3, v_4, v_5\}\rangle$ and $H_2 = \langle\{v_1, v_2, v_3, v_4\}\rangle$. These two subgraphs are isomorphic under the isomorphism $\phi : V(H_1) \to V(H_2)$ defined by

$$\phi : \begin{pmatrix} v_1\, v_3\, v_4\, v_5 \\ v_1\, v_2\, v_4\, v_3 \end{pmatrix}.$$

□

The relationship between a tree's order and size is used to form a new definition of connected graphs containing an arbitrary number of cycles. Recall that for a tree T, $|E(T)| = |V(T)| - 1$. A connected graph of order n and size n is called a *unicyclic* graph. In general, a connected graph G of order n and size $(n-1)+t$ is called a *t-cyclic* graph. Hence, a 1-cyclic graph is a unicyclic graph, a 2-cyclic graph is a *bicyclic* graph and a 3-cyclic graph is a *tricyclic* graph. A t-cyclic graph may contain more than t cycles. This property is shown in Example 1.2.7. The notation $G_{n,t} := \{\text{Order } n \text{ graphs } G |\ G \text{ is } t\text{-cyclic}\}$ is used here on.

Example 1.2.7. *Cycles in a graph.*

Figure 1.10 depicts a 2-cyclic graph containing 3 cycles given by the walks: (u_1, u_2, u_3, u_1), (u_2, u_3, u_4, u_2) and $(u_1, u_2, u_4, u_3, u_1)$.

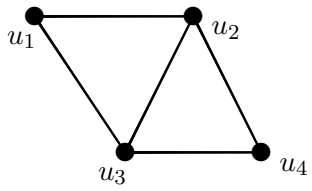


Figure 1.10: A bicyclic graph with 3 cycles.

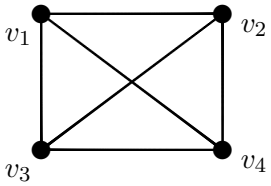


Figure 1.11: A tricyclic graph with 5 cycles.

Figure 1.11 depicts a 3-cyclic graph containing 5 cycles given by: (v_1, v_2, v_3, v_1), (v_1, v_3, v_4, v_1), (v_1, v_2, v_4, v_1), (v_2, v_3, v_4, v_2) and $(v_1, v_2, v_3, v_4, v_1)$. Recall, that due to the four vertices, v_1, v_2, v_3 and v_4, being present across every possible cycle of length 4, these cycles are considered the same. □

A well known Graph Theory result, as found in [5], is stated without proof.

Theorem 1.1. *[5] (The Handshaking Lemma) If G is a graph of size m, then*

$$\sum_{u \in V(G)} deg(u) = 2m.$$

For all other notation and terms not presented here, see [5].

1.3 Graph Domination and Leaf Number

Consider a connected graph G and a set $S \subseteq V(G)$. If every vertex $v \in V(G) - S$ is a neighbour of some vertex in S, then S is called a *dominating set* of G. Let S be a dominating set of G. We call S a *connected dominating set* of G if $\langle S \rangle$ is connected. Note that a graph may have multiple connected dominating sets. The minimum cardinality across all connected dominating sets of G is called the *connected domination number* of G

and is denoted by $\gamma_c(G)$. The *leaf number* of G, denoted $L(G)$, is defined as the number of leaves in a spanning tree of G containing the maximum number of leaves possible.

Example 1.3.1. *Dominating sets and leaf number.*

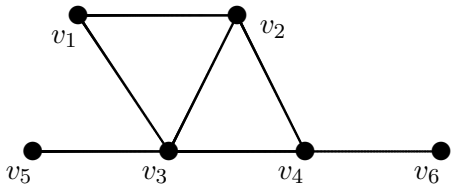


Figure 1.12: A graph G.

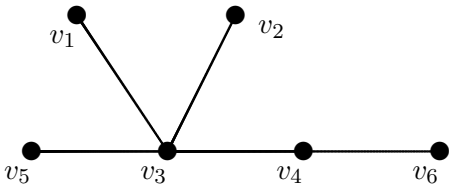


Figure 1.13: A spanning tree T of G.

Shown above in Figure 1.12 is the graph G having $V(G) = \{v_1, v_2, \cdots, v_6\}$ and $E(G) = \{v_1v_2, v_1v_3, v_2v_3, v_2v_4, v_3v_4, v_3v_5, v_4v_6\}$. Figure 1.13 depicts a spanning tree formed by removing edges v_1v_2 and v_2v_4 from G. In both figures, sets $\{v_5, v_1, v_2, v_6\}$ and $\{v_3, v_6\}$ are disconnected dominating sets. Further, $\{v_3, v_2, v_4\}$, $\{v_1, v_2, v_3, v_4\}$, $\{v_3, v_1, v_4\}$ and $\{v_3, v_4\}$ are connected dominating sets. It is clear that $\{v_3, v_4\}$ is a connected dominating set of G having minimum cardinality, thus

$$\gamma_c(G) = |\{v_3, v_4\}| = 2.$$

The leaf number of G is equal to the number of leaves in T, as this is clearly a spanning tree with the maximum number of leaves. This means $L(G) = 4$. □

Note that the vertices of T, in Figure 1.13, that are not leaves are v_3 and v_4, and these are vertices of a minimum cardinality connected dominating set of G, in Figure 1.12. In general,

$$n = L(G) + \gamma_c(G)$$

for all connected graphs G having $|V(G)| = n$. Due to this equality, $\gamma_c(G)$ is studied through $L(G)$.

1.4 Distance Measures

The length of a shortest $u - v$ path in G is the *distance* between a vertex u and a vertex v in G, denoted $d_G(u, v)$, for any connected graph G. When it is clear what graph is being worked with, $d(u, v)$ is used. The distance between the two farthest vertices of G is called the *diameter* of G, denoted by $diam(G)$. Further, define the *distance* (also known as the total distance) of a vertex $v \in V(G)$, denoted $D_G(v)$, to be the sum of all distances between v and every other vertex of G. That is,

$$D_G(v) = \sum_{u \in V(G)} d_G(u, v).$$

When the graph is clear, $D(v)$ is used. The *Wiener index* of a graph G, denoted $W(G)$, is given by

$$W(G) = \sum_{\{u,v\} \subset V(G)} d_G(u, v) = \frac{1}{2} \sum_{v \in V(G)} D_G(v).$$

The *average distance* of graph G is given by

$$\mu(G) = \frac{2}{n(n-1)} \sum_{u \in V(G)} D_G(u) = \binom{n}{2}^{-1} \cdot W(G)$$

where n is the order of G.

Example 1.4.1. *Wiener index and average distance.*

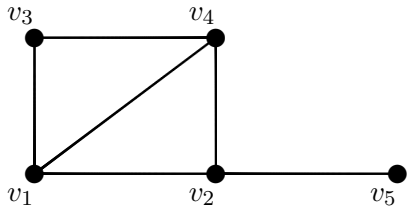


Figure 1.14: Graph G.

Consider graph G in Figure 1.14. First, the distances for all vertices in G is determined as follows:

$$
\begin{aligned}
D(v_1) &= d(v_1, v_2) + d(v_1, v_3) + d(v_1, v_4) + d(v_1, v_5) \\
&= 1 + 1 + 1 + 2 = 5, \\
D(v_2) &= d(v_2, v_1) + d(v_2, v_3) + d(v_2, v_4) + d(v_2, v_5) \\
&= 1 + 2 + 1 + 1 = 5, \\
D(v_3) &= d(v_3, v_1) + d(v_3, v_2) + d(v_3, v_4) + d(v_3, v_5) \\
&= 1 + 2 + 1 + 3 = 7, \\
D(v_4) &= d(v_4, v_1) + d(v_4, v_2) + d(v_4, v_3) + d(v_4, v_5) \\
&= 1 + 1 + 1 + 2 = 5, \text{ and} \\
D(v_5) &= d(v_5, v_1) + d(v_5, v_2) + d(v_5, v_3) + d(v_5, v_4) \\
&= 2 + 1 + 3 + 2 = 8.
\end{aligned}
$$

Second, the Wiener index is determined as follows:

$$
\begin{aligned}
W(G) &= \frac{1}{2}(D(v_1) + D(v_2) + D(v_3) + D(v_4) + D(v_5)) \\
&= \frac{1}{2}(5 + 5 + 7 + 5 + 8) = 15.
\end{aligned}
$$

Last, the average distance:

$$
\begin{aligned}
\mu(G) &= \binom{n}{2}^{-1} \cdot W(G) \\
&= \frac{2}{5 \cdot 4}(15) \\
&= 1.5.
\end{aligned}
$$

□

Needed for Chapter 4 is the Wiener index of a cycle which is stated here without proof as Theorem 1.2.

Theorem 1.2. *[19] The cycle graph C_n, where $n \geq 3$, has*

$$W(C_n) = \begin{cases} \frac{n^3}{8}, & \text{if } n \text{ is even,} \\ \frac{n^3-n}{8}, & \text{if } n \text{ is odd.} \end{cases}$$

1.5 Vertex Unions on Graphs

Let G_1 be a graph with $u_1 \in V(G_1)$ and G_2 be a graph with $u_2 \in V(G_2)$. The *vertex union* of G_1 and G_2, denoted $G_1 \underset{u_1=u_2}{\cup} G_2$, is a new graph G such that u_1 and u_2 become the same vertex, say u, in G. Note that u is a cut-vertex in G, as defined on page 9. Example 1.5.1 depicts the union of graph G_1 and graph G_2 in Figure 1.15 and Figure 1.16.

Example 1.5.1. *Vertex union on two graphs.*

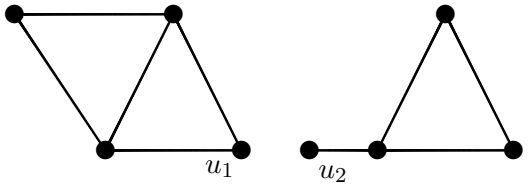


Figure 1.15: Graph G_1 and graph G_2.

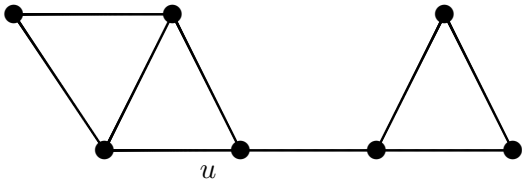


Figure 1.16: A vertex union of G_1 and G_2.

□

1.6 Organic Chemistry and Graph Theory

Flavonoids are a group of bioactive compounds that are extensively found in foodstuffs of plant origin. Their regular consumption is associated with reduced risk of a number of chronic diseases, including cancer, cardiovascular disease (CVD) and neurodegenerative disorders, [12]. Figure 1.17 depicts the molecular form of Quercetin, a flavonoid useful in the treatment of cancer, [20]. Graph Theory cannot readily be used on this compound's molecular form.

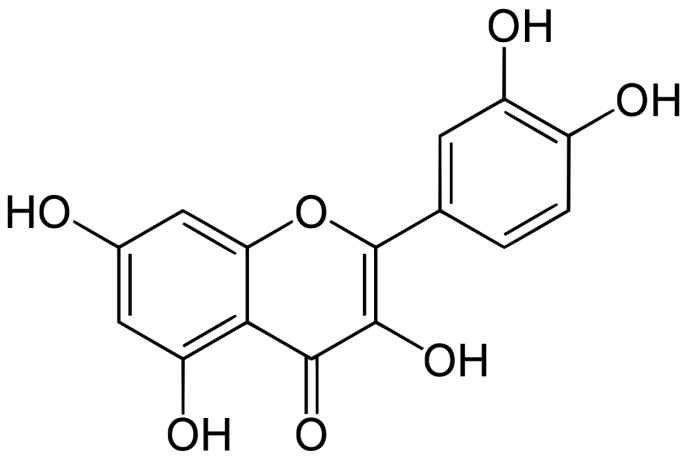


Figure 1.17: The chemical compound Quercetin.

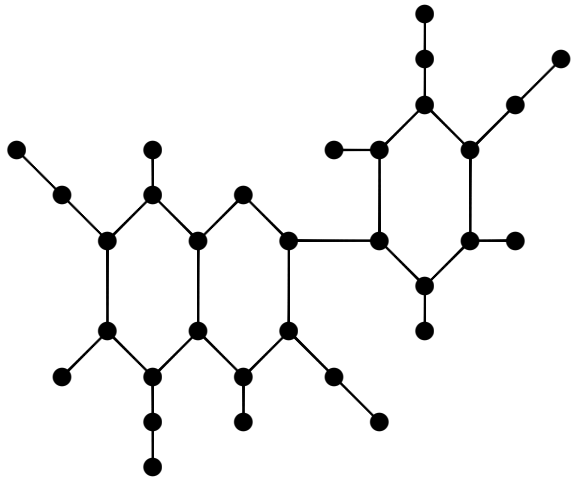


Figure 1.18: The molecular graph for Quercetin.

Therefore, how the molecule is depicted must be transformed so that Graph Theory techniques can then be applied. This new graph, called the *molecular graph*, is obtained by representing each atom by a vertex and each bond by an edge. The molecular graph for Quercetin is shown in Figure 1.18. Note that this representation ignores the types of atoms present in the molecule as well as the presence of triple bonds or double bonds, however, the interested reader may look into use of vertex-weighted graphs or edge-weighted graphs in this regard.

1.7 Overview of Dissertation

The dissertation is divided into 6 chapters of which Chapter 1 defined the necessary definitions and notations used from Graph Theory.

Chapter 2 presents some of the literature relevant to the leaf number, the average distance, and Graph Theory applications in Chemistry.

In Chapter 3, the bounds on the connected domination number of a graph is investigated by studying the bounds on the leaf number. Section 3.2 establishes a lower bound on the leaf number of a graph for general δ while Section 3.3 establishes an upper bound on the leaf number for general δ. Section 3.4 transforms the results of Section 3.2 and Section 3.3 to present bounds on the connected domination number of a graph for general δ.

In Chapter 4, the average distance of unicyclic graphs and tricyclic graphs are investigated. This is done by studying the Wiener index. Section 4.2 presents a theorem that allows us to express the Wiener index of a graph union in terms of the Wiener index of the graphs making up the union. The result is used in various proofs throughout the chapter. Section 4.3 establishes an upper bound and a lower bound on the Wiener index for unicyclic graphs. Section 4.4 establishes an upper bound and a lower bound on the Wiener index for tricyclic graphs. Section 4.5 presents an original result to the Wiener index of t-cyclic graph for any t. Section 4.6 uses the results of Section 4.3 and Section 4.4 to provide bounds on the average distance for both unicyclic graphs and tricyclic graphs.

In Chapter 5, the applicability of modelling compounds' cytotoxicity towards human hepatocarcinoma values using the connected domination number and the average distance of the molecular graph is investigated. Section 5.2 is our hypothesis. Section 5.3 presents the necessary data needed for the model. Section 5.4 presents the model and uses statistical

methods to analyse the output.

Chapter 6 is the closing chapter and brings together all results obtained throughout the dissertation.

1.8 Conclusion

This chapter presents the necessary definitions and notations needed to understand this dissertation. The Graph Theory concepts, leaf number and average distance, described in Section 1.3 and Section 1.4 have been given their own sections as they are the focus of Chapter 3 and Chapter 4, respectively. The vertex union operation described in Section 1.5 is used heavily throughout the proofs of Chapter 4, some of which are original. Finally, the molecular graph of a molecule, as shown in Section 1.6, enables us to use the leaf number and the average distance in an attempt to model biological activity. The next chapter is the literature review.

Chapter 2

Literature Review

2.1 Introduction

This chapter presents a history of the leaf number and the Wiener index of a graph. Special attention is given to the results that build up to the theorems that are proved in later chapters. The chapter is concluded by an overview of the next chapter.

2.2 Leaf Number

In 1979, Sampathkumar and Walikar published [23] following private communication between Walikar and Hedetnieme. This paper defines the connected domination number and then calculates the connected domination number for some of the popular classes of graphs: paths and cycle graphs. Sampathkumar and Walikar also established a few bounds on the graph parameter. It is around this time that the term dominating set was popularised over the other names used, such as "absorbant set", "externally stable set" and "point-point covering set".

Let G be a graph of order n. A result of [23] states that

$$n = L(G) + \gamma_c(G).$$

Thus, the connected domination number of a graph can be investigated by studying the leaf number of the graph. The problem of determining $L(G)$ in graphs of a certain minimum degree was considered independently by several authors. Lovasz and Saks considered this problem in 1987 through a private communication, as mentioned in [11]. Payan, Tchuente, and Xuong in [18] showed that for G a 3-regular graph,

$$L(G) > \frac{n}{4}. \tag{2.1}$$

Independently, Storer, motivated by complexity concerns in [26], showed that for any 3-regular graph G,

$$L(G) \geq \frac{n}{4} + 2. \tag{2.2}$$

The bound of Equation (2.2) was shown, later, to hold for all graphs having minimum degree 3 by Linial and Sturtevant through a private communication, as mentioned in [11]. These previous results were achieved by constructing a spanning tree such that the desired bound was met. The constructions, however, could not be generalised for graphs whose maximum degree does not equal its minimum degree. Kleitman and West in [11] introduced a method of constructing spanning trees known as the *dead leaves method*. This method of spanning tree construction, thoroughly explained in Chapter 3, consists of building a spanning tree from a specially chosen initial subtree. The dead leaves method was used by Kleitman and West in [11] to achieve the same bound as Storer in [26]. The dead leaves method can be generalised to graphs of greater minimum degree. This was implemented in [11] to find two lower bounds on the leaf number given minimum degree. The first result shows that for a graph G having $|V(G)| = n$ and $\delta = 4$,

$$L(G) \geq \frac{2n}{5} + \frac{8}{5}. \tag{2.3}$$

The second result shows that any graph G of order n and minimum degree δ has

$$L(G) \geq \left[1 - b\frac{ln(\delta)}{\delta}\right] n \tag{2.4}$$

where $b > 2.5$ is a constant. Rather than using a set construction method, Inequality (2.4) is proved by showing that some algorithm for the construction of such a tree achieving the bound exists. Griggs and Wu in [7] also obtained the bound in Inequality (2.3), and found that for a graph of order n and $\delta = 5$,

$$L(G) \geq \frac{n}{2} + 2. \tag{2.5}$$

The bounds of Inequality (2.1), Inequality (2.2), Inequality (2.3), Inequality (2.4) and Inequality (2.5) all align with Linial's Conjecture, as mentioned in [7]. The conjecture states that for a graph G having $|V(G)| = n$ and minimum degree δ,

$$L(G) \geq \left(\frac{\delta - 2}{\delta - 1}\right) n + c_\delta \tag{2.6}$$

where c_δ is a constant that depends on δ. By the bounds of Inequality (2.2), Inequality (2.3) and Inequality (2.5), Linial's conjecture is true for graphs with $\delta \leq 5$. Alon in [2], however, proved by studying hypergraphs that for n large enough there exists a graph with minimum degree δ and no dominating set of size less than

$$(1 + o_\delta(1))\frac{1 + ln(\delta + 1)}{\delta + 1} n. \tag{2.7}$$

Since all connected dominating sets are themselves dominating sets, Inequality (2.7) disproves Linial's Conjecture for sufficiently large δ. The same result was obtained later by Caro, West and Yuster in [4]. Our research would indicate that proving Linial's Conjecture false for $\delta \geq 6$ is a difficult problem. Work has been done in proving Linial-like bounds as well as bounds on the leaf number of graphs with forbidden subgraphs. Mukwembi, Munyira and Mafuta in a private communication showed that for a graph G with $|V(G)| = n$

and and minimum degree $\delta \geq 3$,

$$L(G) \geq \frac{\delta - 2}{2\delta - 1} n + \frac{2(\delta + 1)}{2\delta - 1}. \tag{2.8}$$

Showing the bound in Inequality (2.8) to be sharp would be sufficient to disprove Linial's conjecture for larger values of δ. Inequality (2.8) was proved by constructing a spanning tree using the packing method.

The packing method for constructing spanning trees was presented by Dankelmann and Entringer in [6], in an attempt to find spanning trees with the smallest average distance. An upper bound on the leaf number of a graph G having $|V(G)| = n$ was found by Kleitman and West in [11] to be

$$L(G) \leq n - 3 \left\lfloor \frac{n}{\delta + 1} \right\rfloor + 2. \tag{2.9}$$

The proof of Inequality (2.9) considers the set of graphs that have the largest $L(G)$ across all graphs G of minimum degree δ. The authors denote these graphs by $G_{n,\delta}$. The proof fails, however, because $L(G_{n,\delta})$ is calculated without showing that such graphs do indeed have the largest leaf number.

A slightly weaker version of the bound in Inequality (2.2) is proved making use of the dead leaves method and is presented in Chapter 3. The bound of Inequality (2.8) is also proved. The proofs of these two results provide a detailed description of both the dead leaves method and the packing method, and shows off their applicability. A counter example to Inequality (2.9) is presented and we put forward an alternative upper bound on the leaf number is also presented in Chapter 3.

2.3 Wiener Index

In 1947, Harold Wiener introduced the topological graph index in [28] known as the Wiener index for the purpose of determining the approximation formula of the boiling point of

paraffin. This parameter has been shown to have various applications in Chemistry, see for example [3], [21] and [22]. Considerable work has been done to find bounds on the Wiener index of different classes of graphs. Of interest to this dissertation is an upper bound on the Wiener index given the t-cyclic nature of the graph. Wang, Tan and Zhu in [27] showed that the star graph on n vertices with t edges added between any t non-adjacent pairs of vertices has

$$W(G) \geq (n-1)^2 - t. \tag{2.10}$$

This is best possible for any t-cyclic graph. The upper bound for a unicyclic graph has been found in [30] by Yu and Feng to be

$$W(G) \leq \frac{1}{6}(n^3 - 7n + 12). \tag{2.11}$$

Xhanti in [29] found the upper bound for a special class of bicyclic graphs, specifically, cactus graphs. A bicyclic cactus graph G has

$$W(G) \leq \frac{n^3}{6}(n^3 - 13n + 24) - \frac{13}{6}n + 4. \tag{2.12}$$

The *cactus graph* is a special set of connected graphs defined by the nature of their cycles, that is, a connected graph such that any of the contained cycles shares at most one vertex. A *t-cyclic cactus graph* contains exactly t cycles. Gutman, Lie and Wei in [9] showed a cactus graph G with $|V(G)| = n \geq 7$ containing $t \geq 0$ cycles has

$$W(G) \leq \begin{cases} \frac{8}{3}k^3 - 2nk^2 - \frac{2}{3}k + \frac{n^3}{6} - \frac{n}{6} & \text{for } t = 2k, \\ \frac{8}{3}k^3 - 2nk^2 + 6k^2 - 2nk + \frac{16}{3}k + \frac{n^3}{6} - \frac{n}{6} & \text{for } t = 2k+1, \end{cases} \tag{2.13}$$

for some non-negative integer k. Wang, Tan and Zhu in [27] have found the upper bound on the Wiener index of tricyclic graphs, that is,

$$W(G) \leq \frac{1}{6}(n^3 - 19n + 48). \tag{2.14}$$

Soltes in [25] showed that any t-cyclic graph G of order n has

$$W(G) \leq \binom{n}{2} + \sum_{i=1}^{D-1} \left[\binom{n}{2} - m - |S_i(PK_{n,m})| \right], \tag{2.15}$$

where $PK_{n,m}$ is specified in the paper as the path complete graph of size m, order n and having a diameter of D. Further, $S_k(G)$ is defined to be the set of all unordered pairs of non-adjacent vertices in G such that the distance between these vertices does not exceed k. Figure 2.1 depicts the path complete graph on 6 vertices with size 8.

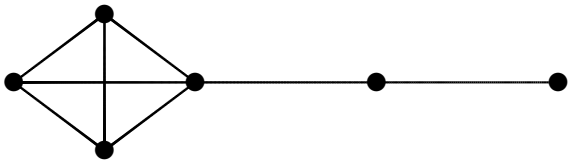


Figure 2.1: Graph $PK_{6,8}$.

Polansky and Bonchev in [19], with the intention of using distance based measures for chemical applications, studied the changes to the Wiener indices of graphs that have had a graph operation applied to them. One of these results is a theorem about the Wiener index of a graph that can be written as a graph union. This result states that if $G = G_1 \underset{u_1=u_2}{\cup} G_2$ with $|V(G_1)| = n_1$ and $|V(G_2)| = n_2$, then

$$W(G) = W(G_1) + W(G_2) + (n_1 - 1)D_{G_2}(u_2) + (n_2 - 1)D_{G_1}(u_1). \tag{2.16}$$

Equation (2.16) is useful to us as this result is used to provide original results to various proofs throughout Chapter 4. Inequality (2.10), Inequality (2.11) as well as Inequality (2.14) are thoroughly examined in Chapter 4 too. An original alternative form to Inequality (2.15) is presented in Section 4.5 and is shown to coincide with the other results.

2.4 Graph Theory and Chemistry

There are problems present in identifying compounds suitable for drug design. Two such obstacles are the large cost to test chemical compounds in the laboratory as well as decid-

ing which compound, from the many possibilities, can be taken to the laboratory for tests.

We believe the second issue posed may be simplified through the modelling of compounds using graphs and Graph Theory techniques to make predictions on the best chemical compounds to test. This approach has been used in papers, such as "Model for prediction of carbonic anhydrase inhibitory activity of sulfonamides", [13]. Various parameters have been used in these models, including the Zagreb Group Parameter, Wiener index and the eccentric connectivity index. To the best of our knowledge, no work has focused on using the connected domination number in a model and we will initiate this study in this dissertation.

2.5 Conclusion

We have seen how the history of the leaf number and the Wiener index developed. In the next chapter we use both the dead leaves method as well as the packing construction method to find lower bounds on the leaf number. This chapter also presents a counterexample to the upper bound on the leaf number presented by Kleitman and West, followed by our own bound that is now best possible.

Chapter 3

Connected Domination Number of Graphs

3.1 Introduction

This chapter establishes bounds on the leaf number of a graph when the order as well as the minimum degree are given. The lower bound on the leaf number is proved for the specific case of $\delta = 3$ and then for the general case of δ using two different methods of forming spanning trees, namely the dead leaves method proposed by Kleitman and West in [11] and the packing method developed by Dankelmann and Entringer [6].

Our result on the upper bound of the leaf number given the minimum degree of the graph closes a non-trivial gap in Theorem 1 of [11]. All graphs used throughout the chapter are necessarily connected.

3.2 Lower Bounds

Recall that for a connected graph G, $L(G)$ denotes the leaf number of G. In this section, lower bounds on $L(G)$, where $|V(G)| = n$ and for various minimum degrees of graph G, is proved. In this regard, first consider $\delta = 3$. The following lemma is a statement about

3-regular graphs. In [11] the following result is stated without proof; our own proof is provided here.

Lemma 3.1. *[11] Let a 3-regular graph, where every edge is on the boundary of a triangle, be G. Then, $G \cong K_4$.*

Proof. Let G be a 3-regular graph having every $e \in E(G)$ being a boundary of a triangle. Let $xy \in E(G)$. By assumption, xy is a boundary of a triangle. Let z be the vertex such that x, y and z form a triangle. Since $deg_G(x) = 3$, vertex x has another neighbor, say x_1. The edge xx_1 must also be on the boundary of a triangle. Next, it is shown that x_1 can only be adjacent to x, y or z.

Suppose, to the contrary, that x_1 has a neighbour $x_2 \neq y$ and $x_2 \neq z$. The edge x_1x_2 must lie on a triangle, say (x_1, x_2, u, x_1). Note that $u \neq y$ and $u \neq z$ since G is 3-regular. It follows that the edge xx_1 cannot be in any other triangle, otherwise x has degree at least 4, a contradiction. Thus, from $deg(x_1) = 3$, x_1 must be joined to both y and z. Since G is 3-regular, $V(H) = \{x, y, z, x_1\}$ and $E(G) = \{xx_1, xy, xz, x_1y, x_1z, yz\}$. Therefore, $G \cong K_4$. □

Before proceeding to Theorem 3.1, a few terms are defined.

The next theorem uses a special method for constructing spanning trees known as the dead leaves method, proposed by Kleitman and West in [11]. The *dead leaves method* is now defined.

Let G be a graph. To construct a spanning tree of G with the desired number of leaves, start with a small subtree of G and then expand the tree step-by-step. Let T denote a subtree with n vertices and l leaves. If x is a leaf of T, then the number of neighbours x has in $G - T$, denoted $d'(x)$, is the *out-degree* of x. Next, expand T by adding the $d'(x)$ edges from x to all the neighbours in $G - T$. This expansion is called an *operation of expansion at* x. A leaf x of T is "dead" if $d'(x) = 0$, that is, no expansion is possible at this leaf.

Hence, x is a leaf in the spanning tree of G.

Let L, D, and N be the number of leaves gained, the number of dead leaves gained and the number of new vertices added respectively. The value of these variables needs to be tracked throughout the construction. To this end, let $f(L, D, N)$ denote a cost function.

Identify an initial subtree T_0 that meets the requirements of Theorem 3.1 below. Thereafter, expand T_0 into a tree of larger order eventually forming the tree T that is a spanning tree. This is achieved by growing the current tree, say T_{i-1}, to T_i ensuring that $f(l, d, n) \geq 0$ for each expansion of T_{i-1}, where l, d and n are calculated when T_{i-1} is expanded to T_i. An expansion satisfying $f(l, d, n) \geq 0$ is called *admissible*. All dead leaves of T_{i-1} are also dead leaves in any subsequent tree. This means that the number of dead leaves gained at stage i are the dead leaves of the tree T_i created by expanding T_{i-1} to T_i. Expanding T_0 using admissible operations results in the final tree being a spanning tree of G, say T, and all leaves are necessarily dead. That is, $L(G) = D$ and the result, Theorem 3.1, follows.

The dead leaves method is now used to prove a lower bound on $L(G)$ for a minimum degree $\delta = 3$ graph G where $|V(G)| = n$ as a result of [11].

Theorem 3.1. *[11] For a graph G of order n and $\delta = 3$,*

$$L(G) \geq \frac{n}{4} + \frac{3}{2}.$$

Proof. Let G be a graph of order n and minimum degree $\delta = 3$. Note that $L(G) \geq n/4+3/2$ if and only if

$$4L(G) - n \geq 6. \tag{3.1}$$

Choose the cost function to be $f = f(L, D, N) = 3L + D - N$. Inequality (3.1) will hold if and only if G contains a spanning tree having $f(L, D, N) \geq 6$.

In this regard, construct a tree T_0 that satisfies $f \geq 6$. Expand T_0 together with subsequent trees to a spanning tree T by a series of expansions where Δn vertices are added such that the number of leaves gained and the number of dead leaves gained are given by Δl and Δn, respectively, at each expansion. Further, $f(\Delta l, \Delta d, \Delta n) \geq 0$ must be satisfied. The final tree T is a spanning tree having all its leaves dead. By definition of $L(G)$, it follows that $3L(G) + L(G) - n \geq 6$. Hence, $f(L, D, N) = 3L + D - N \geq 6$ giving $3L(G) + L(G) = 4L(G) \geq n + 6$, as desired.

Consider now three cases of G whereby a suitable subtree, T_0, is found for each case. Thereafter, admissible operations of expansion is defined to expand T_0 to a spanning tree T of G.

Case 1: There is a $v \in V(G)$ with $deg(v) > 3$.
Let $deg(v) > 3$ and let $\{v_1, v_2, v_3, v_4\} \subseteq N(v)$. Choose T_0 to be the tree with $V(T_0) = \{v, v_1, v_2, v_3, v_4\}$ and $E(T_0) = \{vv_1, vv_2, vv_3, vv_4\}$. Then T_0 is a star with

- $l_0 = 4$,
- $d_0 \geq 0$ and
- $n_0 = 5$.

Hence, $f(l_0, d_0, n_0) = 3l_0 + d_0 - n_0 \geq 3(4) + d_0 - 5 \geq 7$ and so $f(l_0, d_0, n_0) \geq 6$.

Case 2: G is 3-regular and every edge is a boundary of a triangle.
By Lemma 3.1, $G \cong K_4$, and so,

$$L(G) = 3 \geq \frac{n}{4} + \frac{3}{2}.$$

The result then follows.

Case 3: G is 3-regular and at least one of its edges, say e, is not on the boundary of

a triangle.

Let $e = uv$ be one of the edges in G which is not on the boundary of a triangle. Then, $N(u) \cap N(v) = \varnothing$. Further, let u_1, $u_2 \in N(u)$ and v_1, $v_2 \in N(v)$. Choose T_0 to be the tree having $V(T_0) = \{u_2, u_1, u, v, v_1, v_2\}$ and $E(T_0) = \{e, u_2u, u_1u, vv_1, vv_2\}$ as depicted in Figure 3.1. Then,

- $l_0 = 4$,
- $d_0 \geq 0$ and
- $n_0 = 6$.

Hence, $f(l_0, d_0, n_0) = 3l_0 + d_0 - n_0 = 3(4) + d_0 - 6 = 6 + d_0 \geq 6$, as required.

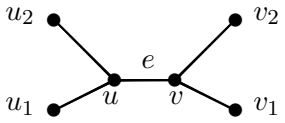


Figure 3.1: Initial subtree of Theorem 3.1, Case 3.

We now introduce three admissible expansions, denoted $O1$ and $O2$, which are sufficient to grow T_0 into T.

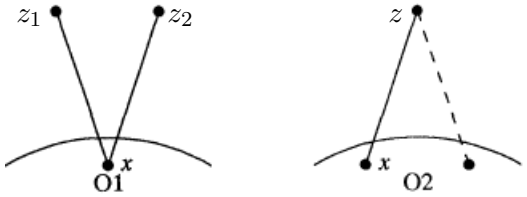


Figure 3.2: Operation $O1$ and Operation $O2$.

Recall that for x to be a leaf in a tree T, $d'(x)$ is the number of neighbours of x that are not yet a part of T. Start off with our initial subtree T_0. Let T_1 be the tree obtained after performing an expansion operation that incorporated more vertices into T_0. After some i expansions there is no more available vertices to add to T_i, that is $T_i = T$ a spanning tree of G. The admissible expansions $O1$ and $O2$ are defined and then shown to be sufficient to build such a tree T_i.

$O1$: Expansion of two or more leaves.

For some current leaf x in T_{i-1} with $d'(x) \geq 2$, let z_1 and z_2 be two of the $d'(x)$ neighbors of x. Expand T_{i-1} at x by joining x to each of z_1 and z_2. Both z_1 and z_2 become leaves but leaf x is lost, so $\Delta l = 2 - 1 = 1$. Note that $\Delta d \geq 0$ and $\Delta n = |\{z_1, z_2\}| = 2$. Therefore, $f(\Delta l, \Delta d, \Delta n) = 3(\Delta l) + \Delta d - \Delta n \geq 3(1) + 0 - 2 = 1$, and so $O1$ is admissible. The expansion is depicted as $O1$ in Figure 3.2.

$O2$: Expansion of a single leaf.

Every current leaf x in T_{i-1} has $d'(x) \leq 1$ and some vertex, say z, outside T_{i-1} has at least 2 neighbours inside T_{i-1}. Expand at x by joining x to z by an edge. Since $d'(x) = 1$, one leaf is added and x is lost as a leaf, $\Delta l = 0$. After expanding at x the two neighbours of z in $T_i = T_{i-1} \cup \{z\}$ become dead giving $\Delta d \geq 2$. Note that $\Delta n = |\{z\}| = 1$. Thus, have $f(\Delta l, \Delta d, \Delta n) = 3(\Delta l) + \Delta d - \Delta n \geq 3(0) + 2 - 1 > 0$, and so $O2$ is admissible.

As $\delta = 3$, any neighbour of a leaf in T_{i-1} that are not in T_{i-1} has either two neighbours in T_{i-1} or two neighbours not in T_{i-1}. Thus, the admissible operations $O1$ and $O2$ is always available until T is spanning. Using expansions $O1$ and $O2$, the initial tree T_0 is expanded to T, with $f(\Delta l, \Delta d, \Delta n) \geq 0$ for each operation. Therefore, the final tree has all of its leaves dead, that is, $D = L(G)$ and $f(L, D, n) = 3L(G) + L(G) - n = 4L(G) - n \geq 6$, completing the proof. □

The proof of Theorem 3.2 can be extended for the case of $\delta = 4$. This is done by choosing a suitable cost function and defining sufficient operations such that a spanning tree with the desired number of leaves may be built. This gives, for a graph G of order n and minimum degree $\delta = 4$,

$$L(G) \geq \frac{2}{5}n + \frac{8}{5}.$$

The lower bound for general δ is found by the work of Mafuta, Mukwembi and Munyira in [15]. First, the following lemma in [15] is stated without proof. Recall the graph inclusion notation defined on page 11.

Lemma 3.2. *[15] Let $T' \subseteq G$ be a subgraph of G that is a tree. There exists a spanning tree T of G such that $T' \subseteq T$ and $L(T) \geq L(T')$.*

Next, the definition of a packing set and how to construct one is presented for a graph G. A *k-packing* set A is a subset of $V(G)$ such that $d_G(x, y) \geq k + 1$ for any $x, y \in A$. A k-packing set B such that for any other k-packing set C where $B \subseteq C$ implies $C = B$ is said to be *maximal*. First, let $A = \{v\}$ for a vertex v of G. Add all vertices u such that $d(u, x) \geq 3$ where $x \in A$ to A. Note that all vertices in G not in A will be within a distance of 2 from A. The following theorem presents the bound on the leaf number given minimum degree and order through the construction of a packing set.

Theorem 3.2 below is of the same nature as Linial's conjecture for the number of leaves in spanning trees. The bound presented below, if sharp, would disprove Linial's conjecture. Theorem 3.2, however, does not examine the sharpness of its bound and this is left as an area of further research.

Theorem 3.2. *[15] Let G be a graph of order n and $\delta \geq 3$. Then,*

$$L(G) \geq \frac{\delta - 2}{2\delta - 1} n + \frac{2(\delta + 1)}{2\delta - 1}.$$

Proof. Let G be a graph of order n and minimum degree δ. Further, suppose there is a maximal 2-packing set, say A, of G. Label each vertex of A by a_i for $i = 1, \cdots, |A|$.

Let T'' be the forest where each component C_i has $V(C_i) = N[a_i]$ and edge set of T'' consists of all edges incident to a_i. Let $l(T'')$ be the number of leaves in T''. It follows that

$$l(T'') = \sum_{v \in A} deg_G(v).$$

There are $|A| - 1$ edges in G such that each of these edges join two distinct vertices of A by the construction of A. Add these edges to T'' to obtain tree T'. Since all $|A| - 1$ edges

are joining two leaves of T'',

$$\begin{aligned} L(T') &\geq L(T'') - 2(|A| - 1) \\ &= \sum_{v \in A} deg_G(v) - 2|A| + 2. \end{aligned} \tag{3.2}$$

Let the vertices of $G - T'$ be B. Then, each $v \in B$ is a distance of one from some $u \in N[A]$. Further, A, $N[A]$ and B form a partition of $V(G)$. Therefore,

$$n = |A| + |N[A]| + |B| = |A| + \sum_{v \in A} deg_G(v) + |B|.$$

It follows that

$$n - |A| - |B| = \sum_{v \in A} deg_G(v),$$

and so, from inequality (3.2),

$$\begin{aligned} L(T') &\geq \sum_{v \in A} deg_G(v) - 2|A| + 2 \\ &= n - |A| - |B| - 2|A| + 2 \\ &= n - 3|A| - |B| + 2. \end{aligned} \tag{3.3}$$

First, consider $|A|$. Due to δ being the minimum degree of all vertices in G,

$$\sum_{v \in A} deg_G(v) \geq \delta|A|$$

and so,

$$\begin{aligned} n &= |A| + \sum_{v \in A} deg_G(v) + |B| \\ &\geq |A| + \delta|A| + |B| \\ &= (\delta + 1)|A| + |B|. \end{aligned}$$

Hence,

$$|A| \leq \frac{n - |B|}{\delta + 1}.$$

By inequality (3.3),

$$\begin{aligned} L(T') &\geq n - 3|A| - |B| + 2 \\ &\geq n - 3\frac{n - |B|}{\delta + 1} - |B| + 2 \\ &\geq \frac{\delta - 2}{\delta + 1}n - \frac{\delta - 2}{\delta + 1}|B| + 2. \end{aligned}$$

Next, the possible values for $|B|$ in terms of δ are determined. If

$$|B| \leq n\frac{\delta - 2}{2\delta - 1} + 2\frac{\delta + 1}{2\delta - 1},$$

then,

$$L(T') \geq \frac{\delta - 2}{2\delta - 1}n + \frac{2(\delta + 1)}{2\delta - 1}.$$

Apply Lemma 3.2 to B, to obtain a spanning tree of G that satisfies the desired bound. If instead

$$|B| > \frac{\delta - 2}{2\delta - 1}n + \frac{2\delta + 1}{2\delta - 1},$$

then T' can be extended to a spanning tree T having a larger leaf number. Since for every vertex v of B there is a vertex v' of $N(A)$ such that v is adjacent to v', let T have

$$E(T') \cup \{vv' | v \in B, v' \in N(A)\}.$$

It follows that

$$L(G) \geq L(T) \geq |B| = \frac{\delta - 2}{2\delta - 1}n + 2\frac{\delta + 1}{2\delta - 1},$$

completing the proof. □

3.3 Upper Bound

The lower bounds on the leaf number given the order and the minimum degree of a graph is found in Section 3.2. This section, for completeness, presents the upper bound on $L(G)$ given $|V(G)| = n$ as well as the minimum degree of graph G. Theorem 3.3 presents an upper bound on $L(G)$ for G given its order and minimum degree of graph G. The result is found in [11].

Theorem 3.3. *[11] Let G be a graph of order n and minimum degree δ. Then,*

$$L(G) \leq n - 3\left\lfloor \frac{n}{\delta + 1} \right\rfloor + 2.$$

As mentioned is Chapter 2, the authors of [11] have not considered all possible graphs in their proof. Example 3.3.1 presents a counterexample to Theorem 3.3.

Example 3.3.1. *Counter example to Theorem 3.3.*
Consider the set of graphs of order 10 and minimum degree 2 graphs. By Theorem 3.3, any such G with $|V(G)| = 10$ and $\delta = 2$ has

$$\begin{aligned} L(G) &\leq n - 3\left\lfloor \frac{n}{\delta + 1} \right\rfloor + 2 \\ &= 10 - 3\left\lfloor \frac{10}{2 + 1} \right\rfloor + 2 \\ &= 3. \end{aligned}$$

Let G' be the graph that is formed by joining a new vertex v to any two vertices, say u and y, of K_9. Then, G' has $|V(G')| = 10$ and $\delta = 2$ so that the above bound in Theorem 3.3 should hold. Form a spanning tree T of G' by removing the edge vy and within K_9, remove all edges not incident to u. Then, T has 9 leaves and so $L(G') \geq L(T) = 9$, a contradiction to the bound stated in Theorem 3.3. □

The coming theorem is a statement of the upper bound on leaf number given order. It is our own result and is now best possible.

Theorem 3.4. *A graph G having order $n \geq 3$ has*

$$L(G) \leq n - 1.$$

Moreover, the bound is best possible for any δ.

Proof. Let a graph G have $|V(G)| = n \geq 3$. As $L(G)$ is defined as the largest leaf number across all spanning trees of G, due to $|E(T)| = |V(T)| - 1$ for any tree T, thus $L(G) \leq n-1$.

The bound's sharpness is now proved. Let H be the graph obtained by joining a new vertex v to δ vertices of K_{n-1}. H has order n and minimum degree δ. Let u be an arbitrary vertex of K_{n-1} that is connected to v. Then, the subgraph T of H with $V(T) = V(H)$ and edge set $\{vv_i \mid v_i \in N(v)\}$, for $i = 1$ to δ, is a spanning tree of H with $n - 1$ leaves. This gives $L(H) \geq n - 1$, and so, $L(H) = n - 1$ from the bound above. □

3.4 Bounds on the Connected Domination Number

Recall that the connected domination number, $\gamma_c(G)$, of graph G is defined as the minimum cardinality across all connected dominating sets of graph G, and

$$\gamma_c(G) + L(G) = n.$$

The following corollary gives bounds to $\gamma_c(G)$ using this relationship, Theorem 3.2 and Theorem 3.4.

Corollary 3.1. *A graph of order n, G, has*

$$1 \leq \gamma_c(G) \leq \frac{\delta + 1}{2\delta - 1}(n - 2).$$

In comparison to Corollary 3.1 the following theorem is presented. According to [10], and the best of our knowledge, Theorem 3.5 is the best known upper bound on the connected domination number given order and minimum degree.

Theorem 3.5. *[10] Let a connected graph, G, have order n and minimum degree $\delta \geq k$. Then,*

$$\gamma_c(G) \leq n\frac{145 + 0.5\sqrt{ln((k+1)} + ln(k+1)}{k+1}.$$

3.5 Conclusion

This chapter established lower bounds and upper bounds on the leaf number. The lower bound on the leaf number was found to be

$$L(G) \geq \frac{n}{4} + \frac{3}{2}$$

for the specific case of $\delta = 3$, and

$$L(G) \geq \frac{2-\delta}{1-2\delta}n - \frac{2(1+\delta)}{1-2\delta}$$

for general δ. A counterexample to Theorem 3.3 in [11] is presented with a correct upper bound on the leaf number given minimum degree, namely

$$L(G) \leq n - 1.$$

The lower bound and upper bound on the Wiener index for both unicyclic graphs and tricyclic graphs is established next in Chapter 4.

Chapter 4

Average Distance in Graphs

4.1 Introduction

This chapter investigates the average distance in unicyclic graphs and tricyclic graphs by studying their Wiener indices. The results of [14] and [9] are used to find the upper bound on $W(G)$ for $G \in G_{n,1}$ and $G \in G_{n,3}$, respectively. Recall that $G_{n,t}$ denotes the set of t-cyclic graphs of order n. These bounds are then transformed into bounds on the average distance for unicyclic graphs and tricyclic graphs. First, a theorem presented by Polansky and Bonchev in [19] that is used to obtain alternative proofs to some of the following results is presented.

The Wiener index bounds for general bicyclic graphs are not studied as no literature has been found on this topic. We present an original proof of a result for the upper bound on average distance for general t-cyclic graphs and we show that the result implies the specific cases of $t = 1$ and $t = 3$, found in [9] and [27], respectively.

4.2 Vertex Unions in Graphs

A result on the Wiener index of a graph, that can be written as a vertex union of two graphs, proved in [19], is presented.

Theorem 4.1. *[19] Let graph G_1 have $|V(G_1)| = n_1$ with $u_1 \in V(G_1)$, and let graph G_2 have $|V(G_2)| = n_2$ with $u_2 \in V(G_2)$. Further, assume that G_1 and G_2 are disjoint and $G = G_1 \underset{u_1=u_2}{\cup} G_2$. Then, $W(G)$ is given by*

$$W(G) = W(G_1) + W(G_2) + D_{G_2}(u_2)(n_1 - 1) + D_{G_1}(u_1)(n_2 - 1).$$

Proof. By definition,

$$\begin{aligned}
W(G) &= \sum_{\{v,u\}\subseteq V(G)} d(v,u) \\
&= \sum_{\{v,u\}\subseteq V(G_1)} d(v,u) + \sum_{\{v,u\}\subseteq V(G_2)} d(v,u) + \sum_{\substack{u\in V(G_1)\setminus\{u_1\} \\ v\in V(G_2)\setminus\{u_2\}}} d(v,u) \\
&= W(G_1) + W(G_2) + \sum_{\substack{u\in V(G_1)\setminus\{u_1\} \\ v\in V(G_2)\setminus\{u_2\}}} [d(u,u_1) + d(u_2,v)] \\
&= W(G_1) + W(G_2) + \sum_{\substack{u\in V(G_1)\setminus\{u_1\} \\ v\in V(G_2)\setminus\{u_2\}}} d(u,u_1) + \sum_{\substack{u\in V(G_1)\setminus\{u_1\} \\ v\in V(G_2)\setminus\{u_2\}}} d(u_2,v) \\
&= W(G_1) + W(G_2) + (n_2 - 1) \sum_{u\in V(G_1)\setminus\{u_1\}} d(u,u_1) + (n_1 - 1) \sum_{v\in V(G_2)\setminus\{u_2\}} d(u_2,v) \\
&= W(G_1) + W(G_2) + D_{G_1}(u_1)(n_2 - 1) + D_{G_2}(u_2)(n_1 - 1).
\end{aligned}$$

This completes the proof. □

4.3 Bounds on the Wiener Index for Unicyclic Graphs

4.3.1 The Lower Bound

The lower bound on $W(G)$ for $G \in G_{n,1}$ is established here. Presented next are two transformations of graphs, stated as processes, and corresponding lemmas about the changes to the Wiener index of these transformed graphs.

Process 4.1. Consider graph X having $v' \in V(X)$, graph Y having $u' \in V(Y)$ and graph H having u, $v \in V(H)$, all three graphs being pairwise disjoint. Identify u with u' and v with v' to obtain graph G. Identify vertices v, v' and u' to obtain the graph G_1^*. Identify vertices u, v' and u' to obtain the graph G_2^*. These graphs are shown in Figure 4.1. □

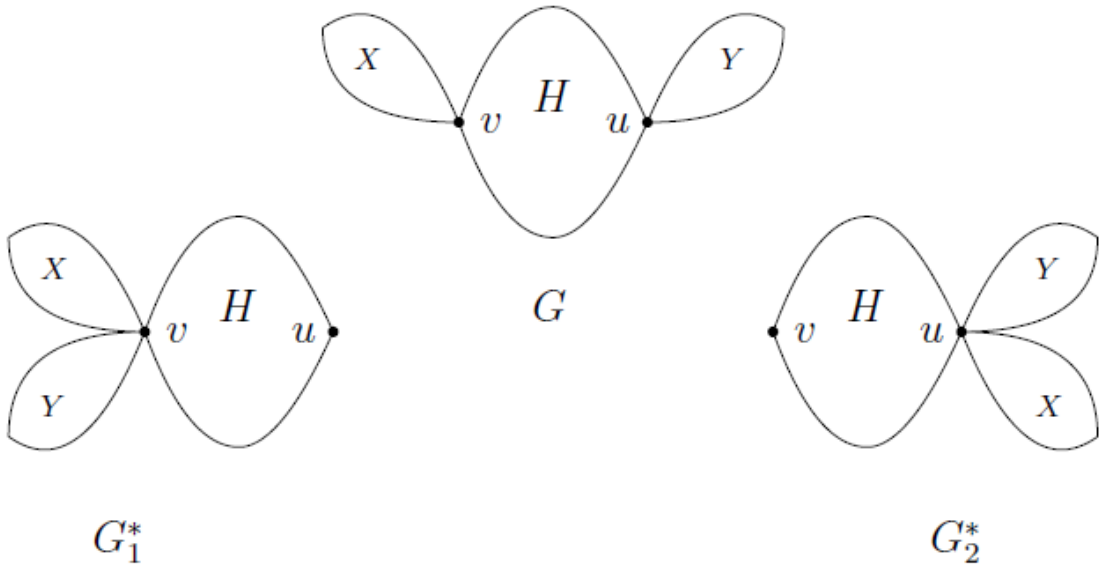


Figure 4.1: Process 4.1 graphs.

Lemma 4.1. *[14] Let the graphs obtained as a result of Process* 4.1 *be* G_1^* *and* G_2^*. *Then,*

$$W(G_1^*) < W(G) \quad or \quad W(G_2^*) < W(G).$$

Proof. Assume that $|V(X)| = n_X$, $|V(Y)| = n_Y$ and $|V(H)| = n_H$. Let $Z = X \underset{v'=v}{\cup} H$ and notice that $G = Z \underset{u=u'}{\cup} Y = ((X \underset{v'=v}{\cup} H) \underset{u=u'}{\cup} Y)$.

By Theorem 4.1,

$$W(Z) = W(X) + W(H) + D_H(v)(n_X - 1) + D_X(v')(n_H - 1)$$

and

$$\begin{aligned} W(G) &= W(Z) + W(Y) + D_Z(u)(n_Y - 1) + D_Y(u')(n_Z - 1) \\ &= W(X) + W(H) + W(Y) + D_H(v)(n_X - 1) + D_X(v')(n_H - 1) \\ &\quad + D_Y(u')(n_X + n_H - 2) + D_Z(u)(n_Y - 1). \end{aligned}$$

Similarly,

$$\begin{aligned} W(G_1^*) &= W(X) + W(H) + W(Y) + D_H(v)(n_X - 1) \\ &\quad + D_X(v')(n_H - 1) + D_Y(u')(n_X + n_H - 2) + D_Z(v)(n_Y - 1). \end{aligned}$$

Therefore,

$$W(G) - W(G_1^*) = (n_Y - 1)[D_Z(u) - D_Z(v)].$$

Now,

$$D_Z(v) = \sum_{x \in V(X)} d_Z(x, v) + \sum_{x \in V(H)} d_Z(x, v)$$

and

$$D_Z(u) = \sum_{x \in V(H)} d_Z(x, u) + \sum_{x \in V(X)} \left[d_X(x, v) + d_H(v, u) \right].$$

Hence, $D_Z(u) > D_Z(v)$ and so $W(G) > W(G_1^*)$. It similarly holds that $W(G) > W(G_2^*)$. This completes the proof. □

Process 4.2. Let a graph which is non-isomorphic to a path having $u \in V(H)$ be H. Consider $C_q = u_0, u_1, \cdots, u_{q-1}, u_0$ for $q \geq 4$. Identify u_0 with u to obtain graph G. Let $G' = G - \{u_{q-1}u_{q-2}\} + \{uu_{q-2}\}$. Figure 4.2 depicts graph G and graph G' below. □

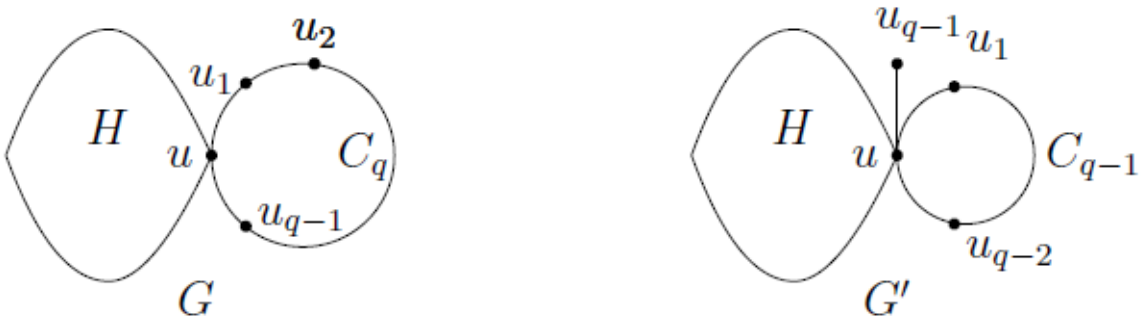


Figure 4.2: Process 4.2 graphs.

Lemma 4.2. *[14] Let the graphs obtained from Process* 4.2 *be* G *and* G'*. Then,*

$$W(G') < W(G).$$

Proof. The Wiener index of G is given by

$$W(G) = \sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=0}^{l-1} d_G(u_i, x) + \sum_{0 \le i < j \le l-1} d_G(u_j, u_i) + \sum_{\{x,y\} \subseteq V(H)} d_G(x, y)$$

and similarly,

$$W(G') = \sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=0}^{l-1} d_{G'}(u_i, x) + \sum_{0 \le i < j \le l-1} d_{G'}(u_j, u_i) + \sum_{\{x,y\} \subseteq V(H)} d_{G'}(x, y).$$

Clearly,

$$\sum_{\{x,y\} \subseteq V(H)} d_G(x, y) = \sum_{\{x,y\} \subseteq V(H)} d_{G'}(x, y).$$

Thus, taking $W(G) - W(G')$ gives,

$$= \sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=0}^{l-1} [d_G(u_i, x) - d_{G'}(u_i, x)] + \sum_{0 \le i < j \le l-1} [d_G(u_j, u_i) - d_{G'}(u_j, u_i)]. \qquad (4.1)$$

Case 1: $l \ge 4$ is even.

The second term of Equation (4.1) is expressible, using Theorem 1.2, as

$$\frac{l^3}{8} - \frac{(l-1)^3 - (l-1)}{8} - d_{G'}(u_{l-1}). \tag{4.2}$$

Note that

$$\begin{aligned} d_{G'}(u_{l-1}) &= \sum_{j=0}^{l-2} d_{G'}(u_{l-1}, u_j) \\ &= 1 + 2\left(2 + 3 + \cdots + \frac{q}{2}\right) \\ &= \frac{l^2 + 2l - 4}{4}. \end{aligned}$$

Thus, (4.2) becomes

$$\frac{l^3}{8} - \frac{(l-1)^3 - (l-1)}{8} - \frac{l^2 + 2l - 4}{4}. \tag{4.3}$$

The first term of Equation (4.1) is equivalent to

$$\begin{aligned} &\sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=0}^{l-1} [d_G(u_0, x) - d_{G'}(u_0, x)] + \sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=0}^{l-1} [d_G(u_i, u_0) - d_{G'}(u_i, u_0)] \\ = &\sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=0}^{l-1} [d_G(u_i, u_0) - d_{G'}(u_i, u_0)]. \end{aligned} \tag{4.4}$$

Now,

$$d_G(u_i, u_0) - d_{G'}(u_i, u_0) = \begin{cases} 0 \text{ if } \quad 0 \le i \le \frac{l-2}{2} \text{ or } i = l-1, \\ 1 \text{ if } \quad \frac{l}{2} \le i \le l-2, \end{cases}$$

and so, Equation (4.4) becomes

$$\begin{aligned} \sum_{x \in V(H)\setminus\{u_0\}} \sum_{i=\frac{l+1}{2}}^{l-2} [d_G(u_i, u_0) - d_{G'}(u_i, u_0)] &= l - 2 - \frac{l-2}{2} \\ &= \frac{l-2}{2}. \end{aligned} \tag{4.5}$$

Combining Equation (4.3) and Equation (4.5), Equation (4.1) becomes

$$\begin{aligned}
&\frac{l-2}{2}+\frac{l^3}{8}-\frac{(l-1)^3-(l-1)}{8}-\frac{l^2+2l-4}{4}\\
=&\frac{l^2-2l}{8}\\
>&0.
\end{aligned}$$

Hence, $W(G) > W(G')$.

Case 2: $l \geq 5$ is odd.

Similar to Equation (4.2), the second term of Equation (4.1) is expressible as

$$\frac{l^3-l}{8}-\frac{(l-1)^3}{8}-d_{G'}(u_{l-1}) \tag{4.6}$$

where

$$\begin{aligned}
d_{G'}(u_{l-1}) &= \sum_{j=0}^{l-2} d_{G'}(u_j, u_{l-1})\\
&= 1+2\left(2+3+\cdots+\frac{l-1}{2}\right)+\frac{(l+1)}{2}\\
&= \frac{l^2+2l-3}{4}.
\end{aligned}$$

Thus, Equation (4.6) becomes

$$\frac{l^3-l}{8}-\frac{(l-1)^3}{8}-\frac{l^2+2l-3}{4}. \tag{4.7}$$

Next, examine the first term of Equation (4.1) which is Equation (4.4):

$$d_G(u_i,u_0)-d_{G'}(u_i,u_0)=\begin{cases}0 & 0\leq i<\frac{l-1}{2} \text{ and } i=l-1,\\ 1 & \frac{l+1}{2}\leq i<l-1,\end{cases}$$

and so, Equation (4.4) becomes

$$\sum_{x\in V(H)\setminus\{u_0\}}\sum_{i=\frac{l+1}{2}}^{l-2}[d_G(u_i,u_0)-d_{G'}(u_i,u_0)] = l-2-\frac{l-1}{2}$$
$$= \frac{l-3}{2}. \tag{4.8}$$

Combining Equation (4.7) and Equation (4.8), Equation (4.1) then becomes

$$\begin{aligned}&\frac{l-3}{2}+\frac{l^3-l}{8}-\frac{(l-1)^3}{8}-\frac{l^2+2l-3}{4}\\ =&\frac{l^2-4l-5}{8}\\ \geq& 0\end{aligned}$$

with equality when $l = 5$. For the next case, where C_5' denotes the modified C_5, by Theorem 4.1,

$$W(G) = W(H) + W(C_5) + D_{C_5}(u_0)(n_H - 1) + D_H(u_0)(l-1)$$

and

$$W(G') = W(H) + W(C_5') + D_{C_5'}(u_0)(n_H - 1) + D_H(u_0)(l-1).$$

Hence,

$$\begin{aligned}W(G) - W(G') &= W(C_5) - W(C_5') + [D_{C_5}(u_0) - D_{C_5'}(u_0)](n_H-1)\\ &= W(C_5) - (W(C_4) + D_{C_5'}(u_4)) + [6-5](n_H-1)\\ &= 15 - 16 + (n_H - 1)\\ &= n_H - 2\\ &> 0.\end{aligned}$$

Since $n_H \geq 3$, the result holds. □

To better understand the next lemma some notation is introduced. We define $G^0(n,t)$ to be a star graph on n vertices, with central vertex v, where t edges have been added such that v is contained in t pair-wise disjoint triangles. This graph is displayed in Figure 4.3.

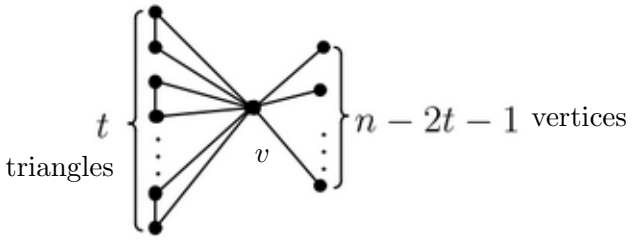


Figure 4.3: Graph $G^0(n,t)$.

Lemma 4.3. *[14] The graph $G^0(n,t)$ has*

$$W(G^0(n,t)) = (n-1)^2 - t.$$

Proof. Consider the graph $G^0(n,t)$. For x, a vertex having $deg(x) = 2$, the distance of x is given by

$$\begin{aligned} D(x) &= 2 + 4(t-1) + 2(n-2t-1) \\ &= 2n - 4. \end{aligned}$$

For u, a vertex having $deg(u) = 1$, the distance of u is given by

$$\begin{aligned} D(u) &= 1 + 4t + 2(n-2t-2) \\ &= 2n - 3 \end{aligned}$$

and, for a vertex v having $deg(v) = n-1$, the distance is given by

$$D(v) = n - 1.$$

The Wiener index of $G^0(n,t)$ is then

$$\begin{aligned} W(G^0(n,t)) &= \frac{1}{2}[(2n-3)(n-2t-1)+2t(2n-4)+(n-1)] \\ &= \frac{1}{2}(2n^2-4n+2-2t) \\ &= (n-1)^2-t. \end{aligned}$$

This finishes the proof. □

Let $G_{n,t}$ denote the set of all t-cyclic graphs of order n. A graph in the set $\mathcal{G}_{n,t} = \{G \in G_{n,t} |$ every block is isomorphic to a path or cycle$\}$ is a *cactus graph*. Figure 4.4 shows a cactus graph on 6 vertices. Note that the graph contains 2 cycles.

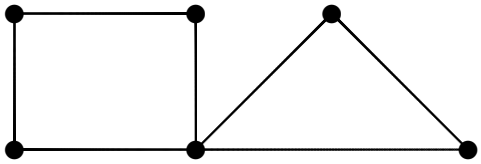


Figure 4.4: A graph in $\mathcal{G}_{6,2}$.

Consider graph G containing b blocks. The graph B having $|V(B)| = b$ where two vertices share an edge given that the corresponding blocks contain a common cut vertex is called the *block graph* of G. The graph B is also a connected graph.

Fact 4.1. Let a graph containing at least 2 cut-vertices be G. Then G has a pair of cut-vertices in one of its blocks.

Proof. Assume a graph G contains b blocks and at least 2 cut-vertices. Further, let B be the block graph of G. Now, $|V(B)| = b$ and let $|E(B)| = m_B$. Note that $m_B \geq b-1$, due to B being connected. We prove the result by contradiction. Suppose for all $v \in V(B)$ that $deg_B(v) \leq 1$. Then,

$$b \geq \sum_{v \in V(B)} deg(v).$$

By the Handshaking Lemma [5],

$$2m_B = \sum_{v \in V(B)} deg(v).$$

Then,

$$2(b-1) \leq 2m_B \leq b$$

resulting in

$$b \leq 2,$$

a contradiction since G has at least 2 cut-vertices implying that $b \geq 3$. The result then follows. □

Theorem 4.2. *[14] If $G \in \mathcal{G}_{n,t}$ has order $n \geq 4$, then*

$$(n-1)^2 - t \leq W(G).$$

Proof. Let V_c be the set of all cut-vertices of G. Pick the graph $G \in \mathcal{G}_{n,t}$ having the minimum $W(G)$. Two properties are proved next.

Property 1. $|V_c| = 1$.

Proof. To the contrary, suppose that $|V_c| > 1$. Let the two vertices in the same block of G, say H, be u and v (Fact 4.1). Denote the component of $G-(V(H)-\{u,v\})$ by X or Y that contains u or v, respectively. The graphs G_1^* and G_2^*, both cactus graphs, are obtained by applying Process 4.1 to G. By Lemma 4.1, $W(G_1^*) < W(G)$ or $W(G_2^*) < W(G)$, a contradiction. Hence, $|V_c| = 1$. ◇

Property 2. $G \cong G^0(n,t)$.

Proof. Assume to the contrary that $G \ncong G^0(n,t)$. All cycles contain a common cut-vertex, say u, by **Property 1**. By repeated applications of Process 4.1, all leaves in G are joined to u. Since $G \ncong G^0(n,t)$, a cycle $C_q = u, u_1, u_1, u_2, ..., u_{q-1}, u$ with $q \geq 4$ exists. Let the graph obtained by applying Process 4.2 to G be $G' = G - \{u_1u_2\} + \{uu_2\}$. By Lemma

4.2, $W(G') < W(G)$, a contradiction and so all cycles must be of length 3. By repeatedly applying Process 4.2, $G \cong G^0(n,t)$. $\diamond$

As a result of Property 1 and Property 2, $W(G^0(n,t)) \leq W(G)$. Finally, due to Lemma 4.3,

$$W(G) \geq (n-1)^2 - t$$

is obtained with equality if and only if $G \cong G^0(n,t)$. $\square$

Corollary 4.1. *If $G \in \mathcal{G}_{n,1}$ has $n \geq 4$, then*

$$W(G) \geq n^2 - 2n.$$

Proof. Let G be a unicyclic graph of order $n \geq 1$ with $t = 1$. By Theorem 4.2,

$$\begin{aligned} W(G) &\geq (n-1)^2 - 1 \\ &= n^2 - 2n. \end{aligned}$$

This completes the proof. $\square$

4.3.2 The Upper Bound

An upper bound on $W(G)$ for $G \in G_{n,1}$ is established in this subsection. The 5 transformation lemmas of [9] are used to prove the result. Each transformation is stated as a process.

Process 4.3. Let the graphs H_i, where $|V(H_i)| = n_i$ for $i = 1, 2, ..., l$, have $x_i \in V(H_i)$. Further, let the cycle graph $C_l = v_1, v_2, \cdots, v_{l-1}, v_l, v_1$ for $l \geq 4$ and let $C_l^{'} = C_l - \{v_l v_1\} + \{v_l v_{l-2}\}$. The graph G is obtained by identifying x_i with v_i from C_l and $H_1, H_2, ..., H_l$. Similarly, G' is obtained. Figure 4.5 depicts graph G and graph G'. $\square$

In the lemma that follows, our own proof is used to prove part (ii).

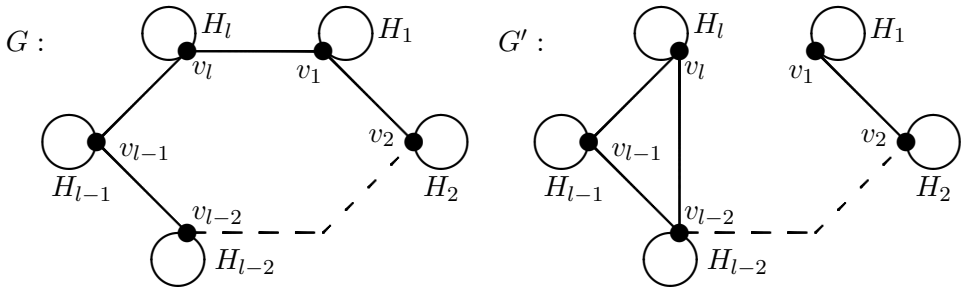


Figure 4.5: Process 4.3 graphs.

Lemma 4.4. *[9] Let the graphs obtained from Process* 4.3 *be* G *and* G'*, with each* n_i *defined as in Process 4.3.*

(i) If $l = 4$*, then* $W(G) \leq W(G')$ *with equality if and only if* $n_1 = n_2$*.*

(ii) If $l \geq 5$*, then* $W(G) < W(G')$*.*

Proof. Part (i) is proved by direct calculation. Consider the sum of the distance of each vertex within each component H_i and then the sum of the distances between each of these components, resulting in

$$\begin{aligned} &W(G') - W(G) \\ &= \sum_{i=1}^{4} \sum_{\{x,y\} \subseteq V(H_i)} [d_{G'}(x,y) - d_G(x,y)] + \sum_{1 \leq i < j \leq 4} \sum_{\substack{x \in V(H_i) \\ y \in V(H_j)}} [d_{G'}(x,y) - d_G(x,y)]. \end{aligned} \tag{4.9}$$

Examining the first term of Equation (4.9), note that $d_G(x,y)$ reduces to $d_{H_i}(x,y)$ and $d_{G'}(x,y)$ reduces to $d_{H_i}(x,y)$. It follows that

$$\begin{aligned} W(G') - W(G) &= \sum_{i=1}^{4} \sum_{\{x,y\} \subseteq V(H_i)} [d_{G'}(x,y) - d_G(x,y)] \\ &= \sum_{i=1}^{4} \sum_{\{x,y\} \subseteq V(H_i)} [d_{H_i}(x,y) - d_{H_i}(x,y)] \\ &= 0. \end{aligned}$$

Thus, Equation (4.9) becomes

$$W(G') - W(G) = \sum_{1 \le i < j \le 4} \sum_{\substack{x \in V(H_i) \\ y \in V(H_j)}} [d_{G'}(x, y) - d_G(x, y)].$$

Since

$$d_G(x, y) = d_{H_i}(x, v_i) + d_G(v_i, v_j) + d_{H_j}(v_j, y),$$

and similarly for $d_{G'}(x, y)$, it follows that

$$\begin{aligned}
&W(G') - W(G) \\
&= \sum_{1 \le i < j \le 4} \sum_{\substack{x \in V(H_i) \\ y \in V(H_j)}} [d_{H_i}(x, v_i) + d_{G'}(v_i, v_j) + d_{H_j}(v_j, y) - d_{H_i}(x, v_i) - d_G(v_i, v_j) - d_{H_j}(v_j, y)] \\
&= \sum_{1 \le i < j \le 4} \sum_{\substack{x \in V(H_i) \\ y \in V(H_j)}} [d_{G'}(v_i, v_j) - d_G(v_i, v_j)] \\
&= \sum_{1 \le i < j \le 4} n_i n_j [d_{G'}(v_i, v_j) - d_G(v_i, v_j)].
\end{aligned}$$

By the construction of G', $d_{G'}(v_i, v_j) - d_G(v_i, v_j) = 0$ for all $1 \le i < j \le 4$, except for the cases of ($i = 1$, $j = 4$) and ($i = 2$, $j = 4$). Figure 4.6 illustrates how the distances are obtained. Thus,

$$\begin{aligned}
W(G') - W(G) &= n_1 n_4 [d_{G'}(v_1, v_4) - d_G(v_1, v_4)] + n_2 n_4 [d_{G'}(v_2, v_4) - d_G(v_4, v_2)] \\
&= n_1 n_4 [2 - 1] + n_2 n_4 [1 - 2] \\
&= n_4 (n_1 - n_2) \\
&\ge 0
\end{aligned}$$

with equality if and only if $n_1 = n_2$.

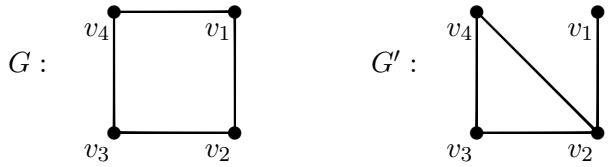


Figure 4.6: Graphs of Case (i).

Next, Part (ii) is proved. To obtain the Wiener index, consider when H_1 is attached to ${C^0}_l$ and then applying Theorem 4.1 to obtain the Wiener index. Then attach H_2 and apply Theorem 4.1 again. Repeat this process until all $H_i's$ are attached and the desired graph is obtained. The notation ${C_l'}^m$ to denote the graph C_l' with H_1 up to H_m attached is used. By this notation, ${C_l'}^0$ is just C_l'. Similarly, ${C_l}^m$ is C_l with H_1 up to H_m attached. Apply Theorem 4.1 to ${C^0}_l$ and H_1 to obtain

$$W(C_l^1) = W({C_l}^0) + W(H_1) + D_{C_l^0}(v_1)(n_{H_1} - 1) + D_{H_1}(v_1)(n_{{C_l}^0} - 1).$$

Apply Theorem 4.1 to C_l^1 and H_2 to obtain

$$\begin{aligned} W(C_l^2) =& W(C_l^1) + W(H_2) + D_{C_l^1}(v_2)(n_{H_2} - 1) + D_{H_2}(v_2)(n_{C_l^1} - 1) \\ =& W({C_l}^0) + W(H_1) + W(H_2) + D_{{C_l}^0}(v_1)(n_{H_1} - 1) \\ &+ (n_{H_2} - 1)D_{C_l^1}(v_2) + D_{H_1}(v_1)(n_{{C_l}^0} - 1) + D_{H_2}(v_2)(n_{C_l^1} - 1). \end{aligned}$$

Apply Theorem 4.1 once more to C_l^2 and H_3 to obtain

$$W(C_l^3) = W({C_l}^0) + \sum_{i=1}^{3} W(H_i) + \sum_{i=1}^{3} D_{C_l^{i-1}}(v_i)(n_{H_i} - 1) + \sum_{i=1}^{3} D_{H_i}(v_i)(n_{{C_l}^{i-1}} - 1).$$

A similar formula holds for $W({C_l'}^3)$. An inductive argument is used to obtain the following Wiener indices:

$$W(G) = W(C_l) + \sum_{i=1}^{l} W(H_i) + \sum_{i=1}^{l} D_{C_l^{i-1}}(v_i)(n_{H_i} - 1) + \sum_{i=1}^{l} D_{H_i}(v_i)(n_{{C_l}^{i-1}} - 1)$$

and

$$W(G') = W(C_l^{'}) + \sum_{i=1}^{l} W(H_i) + \sum_{i=1}^{l} D_{C_l^{'i-1}}(v_i)(n_{H_i} - 1) + \sum_{i=1}^{l} D_{H_i}(v_i)(n_{C_l^{'i-1}} - 1).$$

Subtracting $W(G)$ from $W(G')$ gives

$$W(G') - W(G) = W(C_l^{'}) - W(C_l) + \sum_{i=1}^{l} [D_{C_l^{'i-1}}(v_i) - D_{C_l^{i-1}}(v_i)](n_{H_i} - 1). \quad (4.10)$$

Consider the sigma term of Equation (4.10). Since $(n_{H_i} - 1) \geq 0$ for all i, it follows that

$$\begin{aligned}
&\sum_{i=1}^{l} [D_{C_l^{'i-1}}(v_i) - D_{C_l^{i-1}}(v_i)](n_{H_i} - 1) \\
\geq &\sum_{i=1}^{l} [D_{C_l^{'i-1}}(v_i) - D_{C_l^{i-1}}(v_i)] \\
= &\sum_{i=1}^{l} \left(\sum_{j=i}^{l} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] + \sum_{j=1}^{i-1} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] D_{H_j}(v_j) \right).
\end{aligned}$$

Since $D_{H_j}(v_j)$ is positive for all values of j, it follows that

$$\begin{aligned}
&\sum_{i=1}^{l} \left(\sum_{j=i}^{l} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] + \sum_{j=1}^{i-1} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] D_{H_j}(v_j) \right) \\
\geq &\sum_{i=1}^{l} \left(\sum_{j=i}^{l} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] + \sum_{j=1}^{i-1} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] \right) \\
= &\sum_{i=1}^{l} \sum_{j=i}^{l} [d_{C_l^{'i-1}}(v_i, v_j) - d_{C_l^{i-1}}(v_i, v_j)] \\
= &\sum_{i=1}^{l} [D_{C_l^{'}}(v_i) - D_{C_l}(v_i)] \\
= &2(W(C_l^{'}) - W(C_l)).
\end{aligned}$$

The graph $C_l^{'}$ is $C_3 \cup P_{l-2}$. Apply Theorem 4.1, where any vertex of C_3 is identified with

one of the two leaves of P_{l-2} to obtain

$$\begin{aligned} W(C_l^{'}) &= W(P_{l-2}) + W(C_3) + D_{C_3}(v_{l-2})(l-3) + D_{P_{l-2}}(v_{l-2})(2) \\ &= \frac{1}{6}(l-2)[(l-2)^2 - 1] + 3 + (2)(l-3) + 2\sum_{i=1}^{l-3} i \\ &= \frac{1}{6}(l^3 - 7l + 12). \end{aligned}$$

Recall that

$$W(C_l) = \begin{cases} \frac{l^3}{8} & \text{if } l \text{ is even,} \\ \frac{l^3-l}{8} & \text{if } l \text{ is odd.} \end{cases}$$

Note that

$$\frac{1}{6}(l^3 - 7l + 12) > \frac{l^3}{8}$$

for all $l \geq 4$. Thus, $W(C_l^{'}) > W(C_l)$ holds true for Part (ii). Therefore,

$$\begin{aligned} W(G') - W(G) &= W(C_l^{'}) - W(C_l) + \sum_{i=1}^{l} [D_{C_l{}'i-1}(v_i) - D_{C_l}^{i-1}(v_i)](n_{H_i} - 1) \\ &> 0. \end{aligned}$$

This concludes the proof. □

Process 4.4. Let a graph G_i with $u_i \in V(G_i)$ have order n_i where $n_i \geq 2$ for $i = 1, 2, 3$. Denote a vertex farthest from vertex u_1 in G_1 by v, and a vertex farthest from vertex u_2 in G_2 by w. Identify u_1 with u_2 as the new vertex u to obtain graph G_0. The graphs G, G' and G'' are obtained from G_0 and G_3 by identifying u_3 with u, v and w, respectively. These graphs are shown in Figure 4.7. □

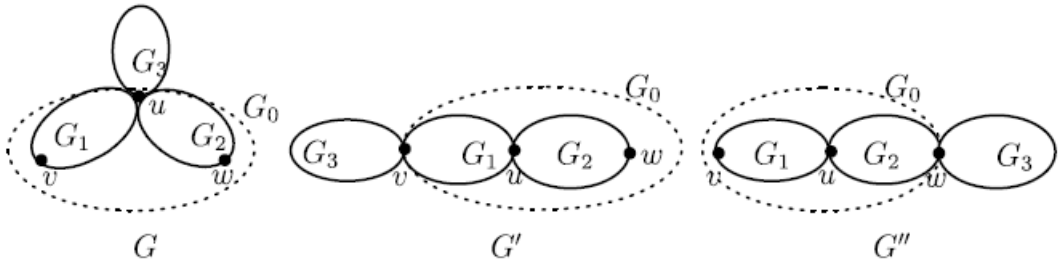


Figure 4.7: Process 4.4 graphs.

Lemma 4.5. *[9] Let the graphs obtained from Process* 4.4 *be G, G' and G''. Then, either*

$$W(G) < W(G') \quad or \quad W(G) < W(G'').$$

Proof. Suppose that $|V(G_0)| = n_0$. By Theorem 4.1,

$$W(G') = W(G_0) + W(G_3) + D_{G_3}(u_3)(n_0 - 1) + D_{G_0}(v)(n_3 - 1)$$

and

$$W(G) = W(G_0) + W(G_3) + D_{G_3}(u_3)(n_0 - 1) + D_{G_0}(u)(n_3 - 1).$$

Subtracting $W(G)$ from $W(G')$, gives

$$\begin{aligned} W(G') - W(G) &= (n_3 - 1)[D_{G_0}(v) - D_{G_0}(u)] \\ &= (n_3 - 1)\left[\sum_{x \in V(G_0)} d_{G_0}(x, v) - \sum_{x \in V(G_0)} d_{G_0}(x, u)\right]. \end{aligned}$$

Now,

$$\sum_{x \in V(G_0)} d_{G_0}(x, v) = \sum_{x \in V(G_1)} d_{G_0}(x, v) + \sum_{x \in V(G_2) \setminus \{u\}} d_{G_0}(x, v)$$

which is similar for the distance of u in G_0. Thus, since $x \in V(G_2)$,

$$d_{G_0}(x,u) + d_{G_0}(u,v) = d_{G_0}(x,v)$$

resulting in

$$\begin{aligned} W(G') - W(G) &= (n_3 - 1)\left[\sum_{x\in V(G_1)} d_{G_0}(x,v) + \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(x,v)\right. \\ &\quad \left. - \sum_{x\in V(G_1)} d_{G_0}(x,u) - \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(x,u)\right] \\ &= (n_3 - 1)\left[\sum_{x\in V(G_1)} d_{G_0}(x,v) - \sum_{x\in V(G_1)} d_{G_0}(x,u) + \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(v,u)\right] \\ &= (n_3 - 1)\left[\sum_{x\in V(G_1)} [d_{G_0}(x,v) - d_{G_0}(x,u)] + \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(v,u)\right]. \end{aligned}$$

For any graph G, $d_G(x,u) + d_G(u,v) \geq d_G(x,v)$, and so

$$\begin{aligned} W(G') - W(G) &\geq (n_3 - 1)\left[-\sum_{x\in V(G_1)} d_{G_0}(v,u) + \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(v,u)\right] \\ &\geq (n_3 - 1)\left[-\sum_{x\in V(G_1)\setminus\{u,v\}} d_{G_0}(v,u) + \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(v,u)\right] \\ &\geq (n_3 - 1)[-(n_1 - 2)d_{G_0}(v,u) + (n_2 - 1)d_{G_0}(v,u)] \\ &= (n_3 - 1)(n_2 - n_1 + 1)d_{G_0}(u,v). \end{aligned} \tag{4.11}$$

By a similar argument for G'', obtain

$$W(G'') - W(G) \geq (n_3 - 1)(n_1 - n_2 + 1)d_{G_0}(u,v). \tag{4.12}$$

If $n_2 \geq n_1$, then $W(G') > W(G)$ by Equation (4.11) and if $n_1 \geq n_2$, then $W(G'') > W(G)$

by Equation (4.12). Therefore, either $W(G') > W(G)$ or $W(G'') > W(G)$, which completes the proof. □

Process 4.5. For $i = 1, 2, 3$, let G_i be a graph with an arbitrarily selected vertex $x_i \in V(G_i)$ and $|V(G_i)| = n_i \geq 2$. Denote a vertex farthest from vertex x_2 in G_2 by u_2. Similarly, x_3 denotes a vertex farthest from u_3 in G_3. The graph G_0 is obtained from G_2, G_3 and $C_3 = v_1, v_2, v_3, v_1$ by identifying x_i with v_i for $i = 2, 3$. Then, the graphs obtained from G_0 and G_1 by identifying x_1 with v_1 and u_2 with u_3, respectively, are G, G' and G'', respectively. Figure 4.8 shows the graphs G_1, G_2, G_3 and G_0. Figure 4.9 depicts the final graphs G, G' and G'' below. □

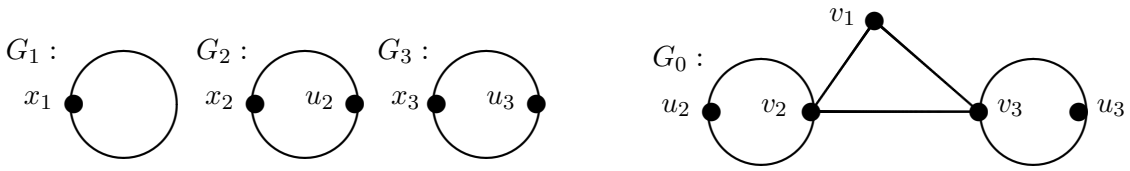


Figure 4.8: Process 4.5 initial graphs.

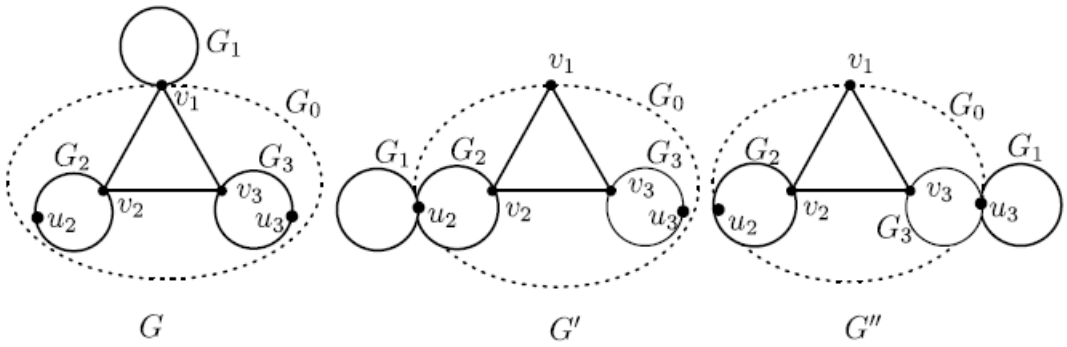


Figure 4.9: Process 4.5 final graphs.

Lemma 4.6. *[9] Let G, G' and G'' be the graphs obtained from Process* 4.5. *Then, either*

$$W(G) < W(G') \quad \textit{or} \quad W(G) < W(G'').$$

Proof. By a similar argument to Lemma 4.5, obtain

$$W(G') - W(G) = (n_1 - 1)[D_{G_0}(u_2) - D_{G_0}(v_1)]. \tag{4.13}$$

Note that since $V(G_0) = V(G_2) \cup V(G_3) \cup \{v_1\}$,

$D_{G_0}(u_2) = \sum_{x \in V(G_2)} d_{G_0}(x, u_2) + \sum_{x \in V(G_3)} d_{G_0}(x, u_2) + d_{G_0}(v_1, u_2)$

$$= \sum_{x \in V(G_2)} d_{G_2}(x, u_2) + \sum_{x \in V(G_3)} [d_{G_3}(x, v_3) + 1 + d_{G_2}(v_2, u_2)] + [1 + d_{G_2}(v_2, u_2)].$$

Similarly, we obtain

$$D_{G_0}(v_1) = \sum_{x \in V(G_2)} [d_{G_2}(x, v_2) + 1] + \sum_{x \in V(G_3)} [d_{G_3}(x, v_3) + 1].$$

Hence, Equation (4.13) becomes

$$\begin{aligned}
&(n_1 - 1) \left[\sum_{x \in V(G_2)} d_{G_2}(x, u_2) + \sum_{x \in V(G_3)} [d_{G_3}(x, v_3) + 1 + d_{G_2}(v_2, u_2)] + [1 + d_{G_2}(v_2, u_2)] \right. \\
&\qquad \left. - \sum_{x \in V(G_2)} [d_{G_2}(x, v_2) + 1] - \sum_{x \in V(G_3)} [d_{G_3}(x, v_3) + 1] \right] \\
=&(n_1 - 1) \left[\sum_{x \in V(G_2)} d_{G_2}(x, u_2) + \sum_{x \in V(G_3)} d_{G_2}(v_2, u_2) + [1 + d_{G_2}(v_2, u_2)] \right. \\
&\qquad \left. - \sum_{x \in V(G_3)} [d_{G_2}(x, v_2) + 1] \right] \\
=&(n_1 - 1) \left[\sum_{x \in V(G_2)} d_{G_2}(x, u_2) + (n_3 + 1) d_{G_2}(v_2, u_2) - (n_2 - 1) - \sum_{x \in V(G_2)} d_{G_2}(x, v_2) \right].
\end{aligned}$$

Making use of the facts that $d_{G_2}(x, u_2) \geq 1$ for any vertex x in $V(G_2) \backslash \{u_2\}$, and

$d_{G_2}(x, v_2) \leq d_{G_2}(v_2, u_2)$. If $n_2 \geq n_3$, then

$$\begin{aligned}
&D_{G_0}(u_2) - D_{G_0}(v_1) \\
&\geq \sum_{x \in V(G_2)} d_{G_2}(x, u_2) - (n_2 - 1) + (n_2 + 1) d_{G_2}(v_2, u_2) - \sum_{x \in V(G_2)} d_{G_2}(x, v_2) \\
&= \sum_{x \in V(G_2) \setminus \{u_2\}} [d_{G_2}(x, u_2) - 1] + \sum_{x \in V(G_2)} [d_{G_2}(v_2, u_2) - d_{G_2}(x, v_2)] + d_{G_2}(v_2, u_2) \\
&> 0.
\end{aligned}$$

Thus, $D_{G_0}(u_2) > D_{G_0}(v_1)$, and so, $W(G') > W(G)$. By a similar argument,

$$\begin{aligned}
&D_{G_0}(u_3) - D_{G_0}(v_1) \\
&\geq \sum_{x \in V(G_3) \setminus \{u_2\}} [d_{G_3}(x, u_3) - 1] \sum_{x \in V(G_3)} [d_{G_3}(v_3, u_3) - d_{G_3}(x, v_3)] + d_{G_3}(v_3, u_3) \\
&> 0.
\end{aligned}$$

This implies that $W(G'') > W(G)$. Thus, either $W(G') > W(G)$ or $W(G'') > W(G)$. □

Process 4.6. Consider any two graphs G_1 and G_2 with $u_1 \in V(G_1)$. Let $|V(G_1)| = n_1 \geq 2$, $|V(G_2)| = n_2 \geq 3$ and assume that G_2 is not a path. Let u and v be the end-vertices of a longest path of G_2. Let G be the graph obtained from G_1, G_2 and a path of order s, say $P_s = p_1, p_2, \cdots, p_s$, by identifying u_1 with u and v with p_1, respectively. Further, let G' be the graph obtained from G_1, G_2 and P_s, by identifying u_1 with p_s and v with p_1, respectively. Finally, let G'' be the graph obtained from G_1, G_2 and P_s by identifying u_1 with p_s and u with p_1, respectively. These graphs are illustrated in Figure 4.10. □

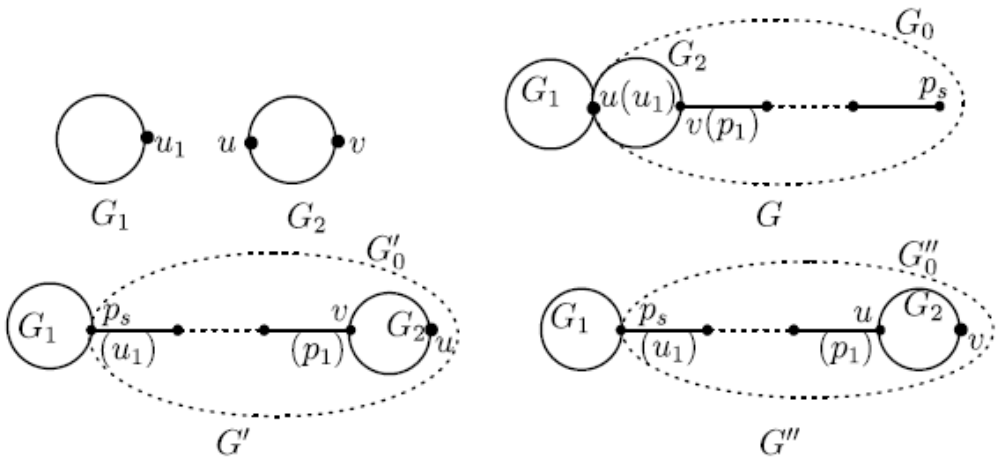


Figure 4.10: Process 4.6 graphs.

Lemma 4.7. *[9] Let the graphs obtained from Process* 4.6 *be G, G' and G''.*

(i) If $D_{G_2}(u) \leq D_{G_2}(v)$, then $W(G) < W(G')$.

(ii) If $D_{G_2}(u) > D_{G_2}(v)$, then $W(G) < W(G'')$.

Proof. Let $G_0 = G - V(G_1)\backslash\{u\}$, $G'_0 = G' - V(G_1)\backslash\{p_s\}$ and $G''_0 = G'' - V(G_1)\backslash\{p_s\}$. Note that $G_0 \cong G'_0$ and $|V(G_0)| = |V(G'_0)| = |V(G''_0)| =: n_0$. By Theorem 4.1,

$$W(G) = W(G_0) + W(G_1) + D_{G_1}(u_1)(n_0 - 1) + D_{G_0}(u)(n_1 - 1),$$
$$W(G') = W(G_0) + W(G_1) + D_{G_1}(u_1)(n_0 - 1) + D_{G_0}(p_s)(n_1 - 1)$$

and

$$W(G'') = W(G''_0) + W(G_1) + D_{G_1}(u_1)(n_0 - 1) + D_{G''_0}(p_s)(n_1 - 1).$$

Consider Part (i) first. Subtracting $W(G)$ from $W(G')$ gives

$$\begin{aligned} W(G') - W(G) &= (n_1 - 1)[D_{G_0}(p_s) - D_{G_0}(u)] \\ &= (n_1 - 1)\left[\sum_{x \in V(G_0)} d_{G_0}(x, p_s) - \sum_{x \in V(G_0)} d_{G_0}(x, u)\right]. \end{aligned} \tag{4.14}$$

Consider the first sigma of Equation (4.14). Write the distance of p_s in G_0 as the distance

from p_s to any vertex in G_2 plus the distance to any vertex in the path excluding v. This is written formally as

$$\sum_{x\in V(G_0)} d_{G_0}(x,v) = \sum_{x\in V(G_2)\setminus\{v\}} [d_{G_0}(x,v) + d_{P_s}(v,p_s)] + \sum_{i=2}^{s} d_{P_s}(p_i,p_s)]$$
$$= D_{G_2}(v) + (n_2-1)(s-1) + D_{P_s}(p_s).$$

Consider the second sigma of Equation (4.14). Write the distance of u in G_0 as the distance from u to any vertex in G_2 plus the distance of u to any vertex in the path graph excluding v. Hence,

$$\sum_{x\in V(G_0)} d_{G_0}(x,u) = \sum_{x\in V(G_2)\setminus\{u\}} d_{G_0}(x,u) + \sum_{i=2}^{s}[d_{P_s}(p_i,v) + d_{G_2}(v,u)]$$
$$= D_{G_2}(u) + D_{P_s}(v) + (s-2)d_{G_2}(v,u).$$

Now,

$$\begin{aligned}
&W(G') - W(G)\\
&= D_{G_2}(v) + (n_2-1)(s-1) + D_{P_s}(p_s) - D_{G_2}(u) - D_{P_s}(v) - (s-2)d_{G_2}(v,u)\\
&= [D_{G_2}(v) - D_{G_2}(u)] + [D_{P_s}(p_s) - D_{P_s}(v)] + (n_2-1)(s-1) - (s-2)d_{G_2}(v,u)\\
&\geq (n_2-1)(s-1) - (s-2)d_{G_2}(v,u).
\end{aligned}$$

By the fact that $D_{P_s}(p_s) = D_{P_s}(v)$, the assumption of $D_{G_2}(u) \leq D_{G_2}(v)$ and due to G_2 being non-isomorphic to a path, $d_{G_2}(v,u) < n_2 - 1$. Hence,

$$W(G') - W(G) \geq (n_1-1)[(n_2-1)(s-1) - (s-2)d_{G_2}(v,u)] > 0.$$

This completes the proof for Part (i).

Consider now Part (ii). Subtracting $W(G)$ from $W(G'')$ gives

$$W(G'') - W(G) = [W(G_0'') - W(G_0)] + (n_1 - 1)[D_{G_0''}(p_s) - D_{G_0}(u)].$$

By Theorem 4.1,

$$W(G_0) = W(G_2) + W(P_s) + D_{P_s}(p_1)(n_2 - 1) + D_{G_2}(v)(s - 1)$$

and

$$W(G_0'') = W(G_2) + W(P_s) + D_{P_s}(p_1)(n_2 - 1) + D_{G_2}(u)(s - 1).$$

Next, consider $W(G_0'') - W(G_0)$. By the hypothesis in Part (ii),

$$W(G_0'') - W(G_0) = (s - 1)[D_{G_2}(u) - D_{G_2}(v)] > 0.$$

Consider now $D_{G_0''}(p_s) - D_{G_0}(u)$ which gives

$$\sum_{x \in V(G_0'')} d_{G_0''}(x, p_s) - \sum_{x \in V(G_0)} d_{G_0}(x, u)$$

$$\begin{aligned}
&= \sum_{i=1}^{s-1} d_{P_s}(p_i, p_s) + \sum_{x \in V(G_2) \setminus \{u\}} [d_{P_s}(p_s, u) + d_{G_0''}(u, x)] \\
&\quad - \sum_{x \in V(G_2) \setminus \{u\}} d_{G_2}(x, u) - \sum_{i=2}^{s} [d_{G_2}(u, v) + d_{P_s}(v, p_i)] \\
&= D_{P_s}(p_s) + (s - 1)(n_2 - 1) + D_{G_2}(u) - D_{G_2}(u) - (s - 1)d_{G_2}(v, u) - D_{P_s}(p_s) \\
&= (s - 1)(n_2 - 1 - d_{G_2}(v, u)).
\end{aligned}$$

It follows that

$$W(G'') - W(G) > (n_1 - 1)(s - 1)(n_2 - 1 - d_{G_2}(v, u)).$$

Since G_2 is non-isomorphic to a path, $d_{G_2}(v,u) < n_2 - 1$. Thus $W(G'') > W(G)$ as required for Part (ii). □

To prove Theorem 4.3 of this subsection, the extremal graph is presented. This graph is called the *saw graph* and is denoted by S_{n,t_1,t_2}. Take P_t and replace each edge with a triangle called a *triangle chain* of length t. To obtain S_{n,t_1,t_2} connect two triangle chains of lengths t_1 and t_2 with the graph P_{n,t_1,t_2}. In what follows it is shown that $S_{n,\lfloor t/2 \rfloor,\lceil t/2 \rceil}$ is the unique graph that has the maximum Wiener index across all unicyclic graphs, when $t = 1$. This graph is shown in Figure 4.11.

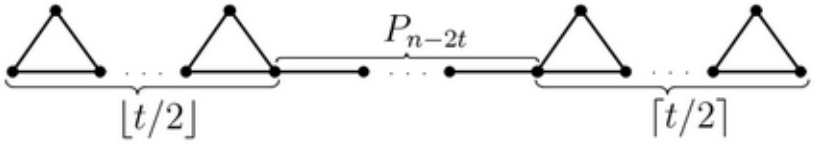


Figure 4.11: Graph $S_{n,\lfloor t \rfloor,\lceil t \rceil}$.

Lemma 4.8. *The saw graph $S_{n,0,1}$ has*

$$W(S_{n,0,1}) = \frac{1}{6}(n^3 - 7n + 12).$$

Proof. Use Theorem 4.1 to obtain

$$\begin{aligned} W(G) &= W(P_{n-2}) + W(C_3) + (D_{C_3}(v))(n-2-1) + (D_{P_{n-2}}(p_1))(3-1) \\ &= \frac{1}{6}(n-2)[(n-2)^2 - 1] + (2)(n-3) + 2\sum_{i=1}^{n-3} i \\ &= \frac{1}{6}(n-2)[(n-2)^2 - 1] + (2)(n-3) + (n-3)(n-2) \\ &= \frac{1}{6}(n^3 - 7n + 12). \end{aligned}$$

This completes the proof. □

Figure 4.12 depicts an example of the saw graph on 7 vertices.

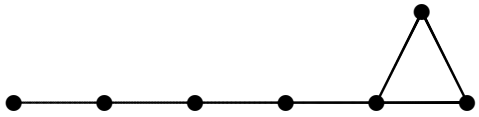


Figure 4.12: $S_{7,0,1}$.

Theorem 4.3. *[9] Let $G \in G_{n,1}$ be of order $n \geq 4$. Then,*

$$W(G) \leq \frac{1}{6}(n^3 - 7n + 12).$$

Equality is met if and only if $G \cong S_{n,0,1}$.

Proof. Assume that G is a unicyclic graph of order n which maximises the Wiener index. Three properties of G which show that $G \cong S_{n,0,1}$ are proved.

Property 1. All cycles have a length of 3.

Proof. Graph G contains cycles of length $l \geq 5$, otherwise Lemma 4.4 is applied to construct a unicyclic graph H where $W(H) > W(G)$, a contradiction. Suppose G contains the cycle $C_4 = v_1, v_2, v_3, v_4$. Let the components of $G - C_4$, containing v_0, v_1, v_2 and v_3, be denoted by H_0, H_1, H_2 and H_3 with $|H_0| \geq |H_i|$ for $i = 1, 2, 3$. If $|H_0| \neq |H_1|$, Lemma 4.4 is used to construct a unicyclic graph G' where $W(G') > W(G)$, a contradiction. If $|H_0| = |H_1| = 1$, then G is the graph C_4. Apply Lemma 4.4 to obtain a unicyclic graph G'' where $W(G'') > W(G)$, a contradiction. Therefore, $|H_0| = |H_1| \geq 2$. Now, G is of the form of the graph in Figure 4.7, and so applying Lemma 4.5 gives another unicyclic graph G''' where $W(G''') > W(G)$, another contradiction. Therefore, G contains cycles of length 3 only. ◇

Property 2. Each block contains up to two cut-vertices. Further, each cut-vertex is contained in at most 2 blocks.

Proof Suppose G has a block containing 3 cut-vertices. Then G is of the form where applying Lemma 4.6 results in a unicyclic graph G' such that $W(G') > W(G)$, a contradiction. Thus, each block contains at most 2 cut-vertices. Suppose a block of G contains a cut-vertex that is shared by three or more blocks. Now, the graph is of the form of G in

Figure 4.7, and so apply Lemma 4.5 to obtain a unicyclic graph G'' where $W(G'') > W(G)$, a contradiction. Therefore each cut-vertex is contained in at most 2 blocks. ◇

Property 3. Graph G contains one pendant path.

Proof. If any block of G, besides the cycle, is not a pendant path, then repeatedly apply Lemma 4.5 to obtain the contradiction that a unicyclic graph G' where $W(G') > W(G)$ exists. As the cycle is of length 3, G contains at most 3 pendant paths. By the repeated application of Lemma 4.7, a unicyclic graph G'' such that $W(G'') > W(G)$ exists where only one pendant path is contained in G'', a contradiction. Thus, G contains one pendant path only. ◇

By Property 1, Property 2 and Property 3, $G \cong S_{n,0,1}$. Apply Lemma 4.8 to obtain $W(G) = \frac{1}{6}(n^3 - 7n + 12)$. Therefore, for any unicyclic graph G',

$$W(G') \leq \frac{1}{6}(n^3 - 7n + 12),$$

completing the proof. □

4.4 Bounds on the Wiener Index for Tricyclic Graphs

4.4.1 The Lower Bound

This subsection establishes a lower bound on $W(G)$ for $G \in G_{n,3}$. Our proof is based on the following transformation lemmas developed in [27]. A tricyclic graph may contain 3, 4, 6 or 7 cycles. It follows that $G_{n,3} = {}^3\mathrm{G}_{n,3} \cup {}^4\mathrm{G}_{n,3} \cup {}^6\mathrm{G}_{n,3} \cup {}^7\mathrm{G}_{n,3}$, where ${}^r\mathrm{G}_{n,3}$ is the set of all $G \in G_{n,3}$ containing r cycles, for $r = 3, 4, 6, 7$. Shown in Figure 4.13 is an example of a tricyclic graph containing 4 cycles. Note that a tricyclic graph cannot contain 5 cycles, see [27].

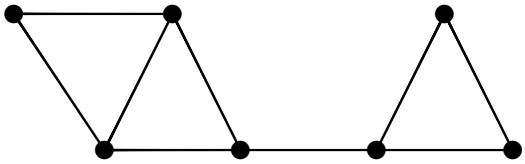


Figure 4.13: A graph in ${}^4\mathrm{G}_{n,3}$.

Let $\mathcal{M}$ be the set $\{G \in G_{n,3} | diam(G) = 2\}$. This set contains the graphs M_1 to M_{12} and are shown in Figure 4.14.

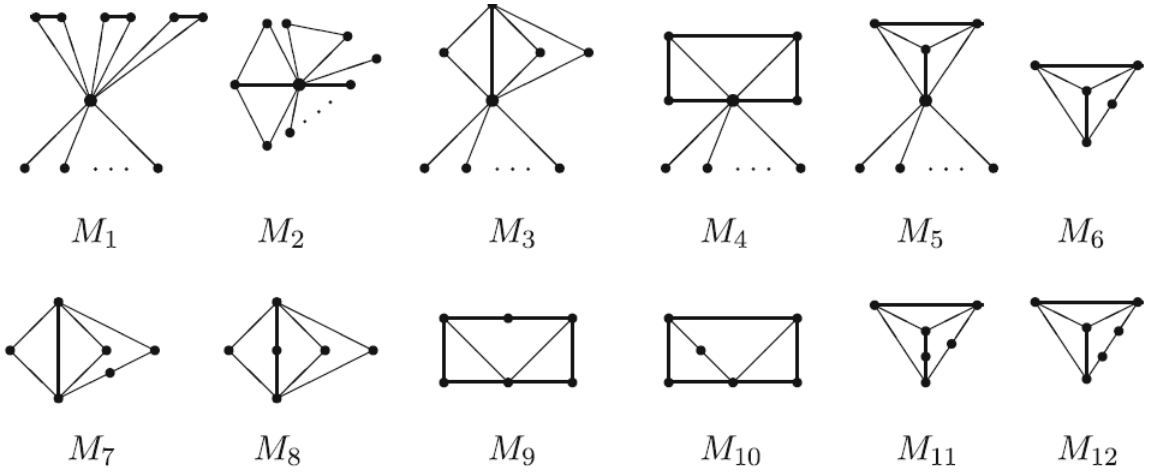


Figure 4.14: The set $\mathcal{M}$.

Theorem 4.4. *[27] If $G \in G_{n,3}$, then*

$$W(G) \geq (n-1)^2 - 3$$

with equality if and only if $G \in \mathcal{M}$.

Proof. Let G be a tricyclic graph having the minimum Wiener index. Consider the form that G can take. By the definition of tricyclic graphs, $|E(G)| = n + 2$. Let the number of pairs of vertices at least a distance of 2 apart be D. Then, since each edge of G connects

vertices of a pair at a distance of 1,

$$\begin{aligned} D &= \binom{n}{2} - (n+2) \\ &= \frac{1}{2}(n^2 - n) - (n+2). \end{aligned}$$

Hence,

$$\begin{aligned} W(G) &\geq (n+2) + 2D \\ &= (n+2) + (n^2 - n - 2n - 4) \\ &= n^2 - 2n - 2 \\ &= (n-1)^2 - 3. \end{aligned}$$

Therefore, for any tricyclic G, $W(G) \geq (n-1)^2 - 3$ and equality is met if and only if $G \in \mathcal{M}$, that is, $\text{diam}(G) = 2$. □

This method of proof is robust. The bound obtained through direct calculation in Theorem 4.2 is obtained by setting the size of G to n in the above proof.

4.4.2 The Upper Bound

An upper bound on $W(G)$ for $G \in G_{n,3}$ is established in this subsection. The results presented in [27] are used. Let a tricyclic graph that maximises the Wiener index be denoted by $J(n)$. This graph is formed by identifying one vertex of K_4 with one of the leaves of the path graph on $n-3$ vertices. This graph on 6 vertices is depicted in Figure 4.15.

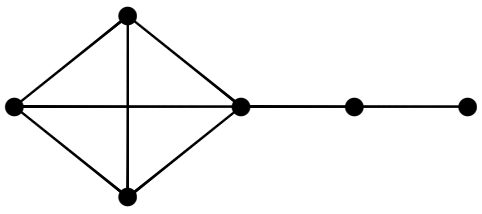


Figure 4.15: $J(6)$.

Note that $D_{J(n)}(v) = \frac{1}{2}(n^2-n-6)$ where v is the only leaf of $J(n)$. In the following lemma an original proof to [27] is used. Let G be a graph containing a vertex v. The *eccentricity* of v is the largest distance between v and any other vertex of G.

Lemma 4.9. *[27] Let $G \in G_{n,3}$ and let v be a vertex in G. Then,*

$$D_G(v) \leq \frac{1}{2}(n^2 - n - 6).$$

Equality is met if and only if $G \cong J(n)$, v is the only leaf in G, $n \geq 6$.

Proof. Let $G \in G_{n,3}$ for $n \geq 6$ and let u be a vertex of G having the largest eccentricity. Now,

$$D(u) = 1 \cdot n_1 + 2 \cdot n_2 + \cdots + e \cdot n_e,$$

where the eccentricity of u is e and n_i denotes the number of vertices a distance of i from vertex u for $1 \leq i \leq e$. The largest e can be is $n-3$ since G is tricyclic. For $D(u)$ to be as large as possible, n_1, n_2, $\cdots$, n_{e-1} must be as small as possible. Due to G being connected, the smallest this can be is 1, that is, take $n_1 b = n_2 = \cdots = n_{e-1} = 1$. This means $n_e = 3$ since $n_1 + n_2 + \cdots + n_{e-1} + n_e = n - 1$. Note that this occurs only for $G \cong J(n)$ and u is the unique leaf. Now,

$$\begin{aligned} D(u) &= 1 \cdot n_1 + 2 \cdot n_2 + \cdots + e \cdot n_e \\ &\leq 1 + 2 + \cdots + 3(n-3) \\ &= \sum_{i=1}^{n-4} i + 3(n-3) \\ &= \frac{1}{2}(n-4)(n-3) + 3(n-3) \\ &= \frac{1}{2}(n^2 - n - 6). \end{aligned}$$

Equality is met if and only if $G \cong J(n)$ and u is the only leaf of G. □

Lemma 4.10. *[27]* $W(J(n)) = \frac{1}{6}(n^3 - 19n + 48)$.

Proof. By Theorem 4.1,

$$\begin{aligned} W(J(n)) &= W(K_4) + W(P_{n-3}) + 3D_{P_{n-3}}(u) + (n-4)D_{K_4}(u) \\ &= 6 + \frac{1}{6}(n-3)[(n-3)^2 - 1] + 3\sum_{i=1}^{n-4} i + 3n - 12 \\ &= \frac{1}{6}(n^3 - 9n^2 + 26n + 12) + \frac{1}{2}(3n^2 - 15n + 12) \\ &= \frac{1}{6}(n^3 - 19n + 48). \end{aligned}$$

This completes the proof. □

In the proof for the next theorem, an original proof to Case 1 is provided.

Theorem 4.5. *[27] For any $G \in G_{n,3}$ of order $n \geq 5$,*

$$W(G) \leq W(J(n))$$

with equality if and only if $G \cong J(n)$ and $n \geq 6$.

Proof. Let a tricyclic graph containing vertex v with neighbour u be G.

Case 1: v is a leaf

Proceed by induction. Note that since G contains a leaf, $G = G' \underset{u_1=u_2}{\cup} P_2$. For the base case of $n = 5$, $G \in \{M_i : 3 \leq i \leq 6, n = 5\}$ and $J(5) \cong M_5$. By direct calculation and Lemma 2.4.2, $W(M_i) = 13 = W(J(5))$ for $i = 3, 4, 5, 6$. Suppose that $W(G) \leq W(J(n))$ for all $n \leq k$. Then, for $n = k + 1$, apply Theorem 4.1 to G and $J(n)$ to obtain

$$W(G) = W(G') + W(P_2) + (k-1)D_{P_2}(u) + (2-1)D_{G'}(u)$$

and

$$W(J(k+1)) = W(J(k)) + W(P_2) + (k-1)D_{P_2}(u) + (2-1)D_{J(k)}(u).$$

Note that graph G' is tricyclic and so $W(G') \leq W(J(k))$ by the inductive assumption.

Further, $D_{G'}(u) \leq D_{J(k)}(u)$ by Lemma 4.9. Thus, the result holds for Case 1.

Case 2: G has no leaf

By Lemma 4.9, $D_G(v) \leq \frac{1}{2}(n^2 - n - 6)$ for all $v \in V(G)$. Then,

$$\begin{aligned} W(G) &= \frac{1}{2} \sum_{v \in V(G)} D_G(v) \\ &< \frac{1}{4}(n^2 - n - 6) \times n \\ &< \frac{1}{6}(n^3 - 19n + 48) \\ &= W(J(n)) \end{aligned}$$

for all $n \geq 1$. Thus, the result holds for Case 2.

Note that for the case of $n = 5$, each graph in $\{M_i : 3 \leq i \leq 6, n = 5\}$ has the maximum Wiener index. Thus, the tricyclic graph having maximum Wiener index is isomorphic to $J(n)$ for $n \geq 6$, completing the proof. □

4.5 Bounds on the Wiener Index for t-Cyclic Graphs

This section presents the extremal graph with regards to the Wiener index of a t-cyclic graph given t in Conjecture 4.6. We show that Conjecture 4.6 implies the results of Theorem 4.3 and Theorem 4.5 presented earlier in the chapter.

A method of construction to obtain the extremal graph on n vertices that is t-cyclic which we denote by $P_{n,t}$ is described below.

The desired graph is constructed using a path $P_n = v_1, v_2, \cdots, v_n$ as a base to which the desired number of t edges is added. First, join vertex v_1 to v_3 by an edge. Next, join v_2 to v_4, followed by forming the edge v_1v_4. Repeatedly joining edges in this way gives graph

having $n-1+t$ edges and is denoted $P_{n,t}$. The 5 graphs across Figure 4.16 to Figure 4.20 depicts the formation of $P_{6,4}$.

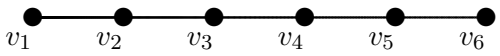


Figure 4.16: Graph P_6.

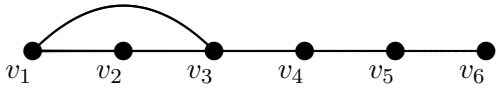


Figure 4.17: Graph $P_{6,1}$.

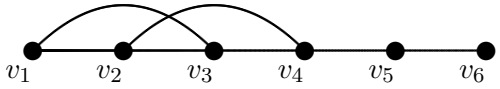


Figure 4.18: Graph $P_{6,2}$.

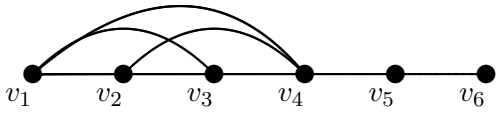


Figure 4.19: Graph $P_{6,3}$.

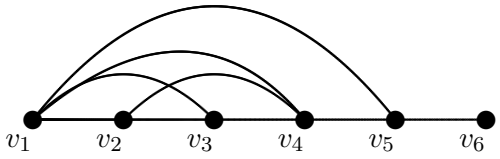


Figure 4.20: Graph $P_{6,4}$.

Conjecture 4.6. *A t-cyclic graph G on n vertices has*

$$W(G) \leq \frac{1}{6}(n^3 - n) + t(j + 1 - n) - \frac{1}{6}(j - 1)j(j + 1),$$

where $j = \left\lceil \frac{-1+\sqrt{(1+8t)}}{2} \right\rceil$. *Equality is met if and only if* $G \cong P_{n,t}$.

A unicyclic graph G has $t = 1$. Theorem 4.6 states that $j = \left\lceil \frac{-1+\sqrt{(1+8(1))}}{2} \right\rceil = 1$ and so,

$$\begin{aligned} W(G) &\leq \frac{1}{6}(n^3 - n) + 1(1 - n + 1) - \frac{0 \times 1 \times 2}{6} \\ &\leq \frac{1}{6}(n^3 - n) - n + 2 \\ &= \frac{1}{6}(n^3 - 7n + 12). \end{aligned}$$

This is exactly the result obtained in Theorem 4.3. A tricyclic graph G has $t = 3$. Theorem 4.6 states that $j = \left\lceil \frac{-1+\sqrt{(1+8(3))}}{2} \right\rceil = 2$ and so,

$$\begin{aligned} W(G) &\leq \frac{1}{6}(n^3 - n) + 3(2 - n + 1) - \frac{1 \times 2 \times 3}{6} \\ &\leq \frac{1}{6}(n^3 - n) - 3n + 8 \\ &\leq \frac{1}{6}(n^3 - 19n + 48). \end{aligned}$$

Again, this is exactly the result obtained in Theorem 4.5.

4.6 Bounds on Average Distance

4.6.1 Unicyclic Graphs

From Corollary 4.1 and Theorem 4.3 the following corollary is obtained.

Corollary 4.7. *For a unicyclic graph* G *of order* $n \geq 4$,

$$2 \cdot \frac{n-2}{n-1} \leq \mu(G) \leq \frac{1}{3} \cdot \frac{n^3 - 7n + 12}{n(n-1)}.$$

Proof. By Corollary 4.1,

$$(n-1)^2 - 1 = n^2 - 2n \leq W(G).$$

Thus, the average distance of this bound is

$$\begin{aligned}\mu(G) &\geq \frac{(n-1)^2-1}{\binom{n}{2}} \\ &= 2 \cdot \frac{n-2}{n-1}.\end{aligned}$$

Similarly, by Theorem 4.5,

$$\begin{aligned}\mu(G) &\leq \frac{\frac{1}{6}(n^3-7n+12)}{\binom{n}{2}} \\ &= \frac{1}{3} \cdot \frac{n^3-7n+12}{n(n-1)}.\end{aligned}$$

This completes the proof. □

4.6.2 Tricyclic Graphs

From Theorem 4.4 and Theorem 4.5, the following corollary is obtained.

Corollary 4.8. *For a tricyclic graph G of order $n \geq 6$,*

$$2 \cdot \frac{n^2-2n+2}{n(n-1)} \leq \mu(G) \leq \frac{1}{3} \cdot \frac{n^3-19n+48}{n(n-1)}.$$

Proof By Theorem 4.4,

$$\begin{aligned}\mu(G) &\geq \frac{(n-1)^2-3}{\binom{n}{2}} \\ &= 2 \cdot \frac{n^2-2n+2}{n(n-1)}.\end{aligned}$$

Further, by Theorem 4.5,

$$\begin{aligned}\mu(G) &\leq \frac{\frac{1}{6}(n^3-19n+48)}{\binom{n}{2}} \\ &= \frac{1}{3} \cdot \frac{n^3-19n+48}{n(n-1)}.\end{aligned}$$

This completes the proof. □

4.7 Conclusion

We proved, in Theorem 4.1, that for a graph $G = G_1 \underset{u_1=u_2}{\cup} G_2$,

$$W(G) = W(G_1) + W(G_2) + D_{G_2}(u_2)(|V(G_1)| - 1) + D_{G_1}(u_1)(|V(G_2)| - 1).$$

This theorem is used throughout the chapter to prove the various Wiener index bounds, some of which are our own proofs. The graph $G^0(n, t)$ is defined and shown to be the graph having the smallest Wiener index across all t-cyclic graphs. Further, Theorem 4.2 shows that for a t-cyclic graph G where $|V(G)| = n \geq 4$,

$$W(G) \geq (n-1)^2 - t,$$

with equality if and only if $G \cong G^0(n, t)$. The extremal graph using graph transformations is also specified. A combinatorial approach is used to identify $G^0(n, 3)$. Theorem 4.3 states that a unicyclic graph G of order $n \geq 4$ has

$$W(G) \leq \frac{1}{6}(n^3 - 7n + 12),$$

with equality for $G \cong S_{n,0,1}$. Theorem 4.5 states that, G, a tricyclic graph of order $n \geq 6$,

$$W(G) \leq \frac{1}{6}(n^3 - 19n + 48),$$

with equality for $G \cong J(n)$.

The lower bound and upper bound on the Wiener index for both unicyclic graphs and tricyclic graphs are then used to obtain bounds on the average distance for unicyclic graphs and tricyclic graphs, respectively. We also present the t-cyclic graph having the largest

Wiener index for a given order and t. In the next chapter the leaf number and the average distance are used as model parameters and their suitability is tested.

Chapter 5

Average Distance and Leaf Number as Predictors for Biological Properties

5.1 Introduction

In [28], Harold Wiener used the Wiener index as a single parameter to predict boiling points of compounds. In [16], Mukwembi and Nyabadza used conduction of a graph as a single parameter to predict boiling points of compounds while in [17] they used graph activity as a single parameter to predict the efficacy of compounds in treating melanoma cancer. In this chapter, we explore if leaf number or average distance can be useful as a single parameter to predict the efficacy of compounds, specifically flavonoids, in treating hepatocellular incarconoma.

The HepG2 cell line is a group of cells extracted from an individual suffering from hepatocellular carcinoma (HCC), a common type of liver cancer. These cells can be used for drug studies targeting liver cancer. This is what was done by Ahmed, Moawad, Owis, Abouzid and Ahm in [1]. The authors extracted certain flavoinoids from the *Calligonum*

polygonoides plant and effectiveness of the flavonoids in inhibiting the growth of the HepG2 cells were calculated and tabulated. The concentration of each compound required to inhibit the cell growth by 50% is defined to be the IC_{50} value and is the measure of compound efficacy. It is these tabulated IC_{50} values for each compound that are then modelled using the leaf number and the average distance for these compounds' molecular graphs. The output is then analysed.

Shown in Figure 5.1, on the next page, are the 14 compounds that were extracted in [1]. Each oxygen atom is represented by an O, an H represents a hydrogen atom and a C represents a carbon atom. Each cycle of length 6 has a carbon atom present but these are omitted for clarity. These carbons exist at each angle in the cycles. A line between two atoms represents a single bond and a double line a double bond. The bolder lines and dashed lines are to be interpreted as a single bond.

When two or more atoms are written next to each other, for example $COOC_4H_9$, this means that the molecule specified by these atoms is bonded. There are hydrogen atoms that are not shown but are known to be present as the valence of the atoms need to be met. When a C is bonded to an OH and there is a neighbour to another C also bonded to a OH, we call this neighbouring *Hydroxyls*.

5.2 Hypothesis

The null hypothesis is that a model using average distance as a parameter is better at replicating the dataset compared to the leaf number. This is believed due to the application of the Wiener index in Chemistry.

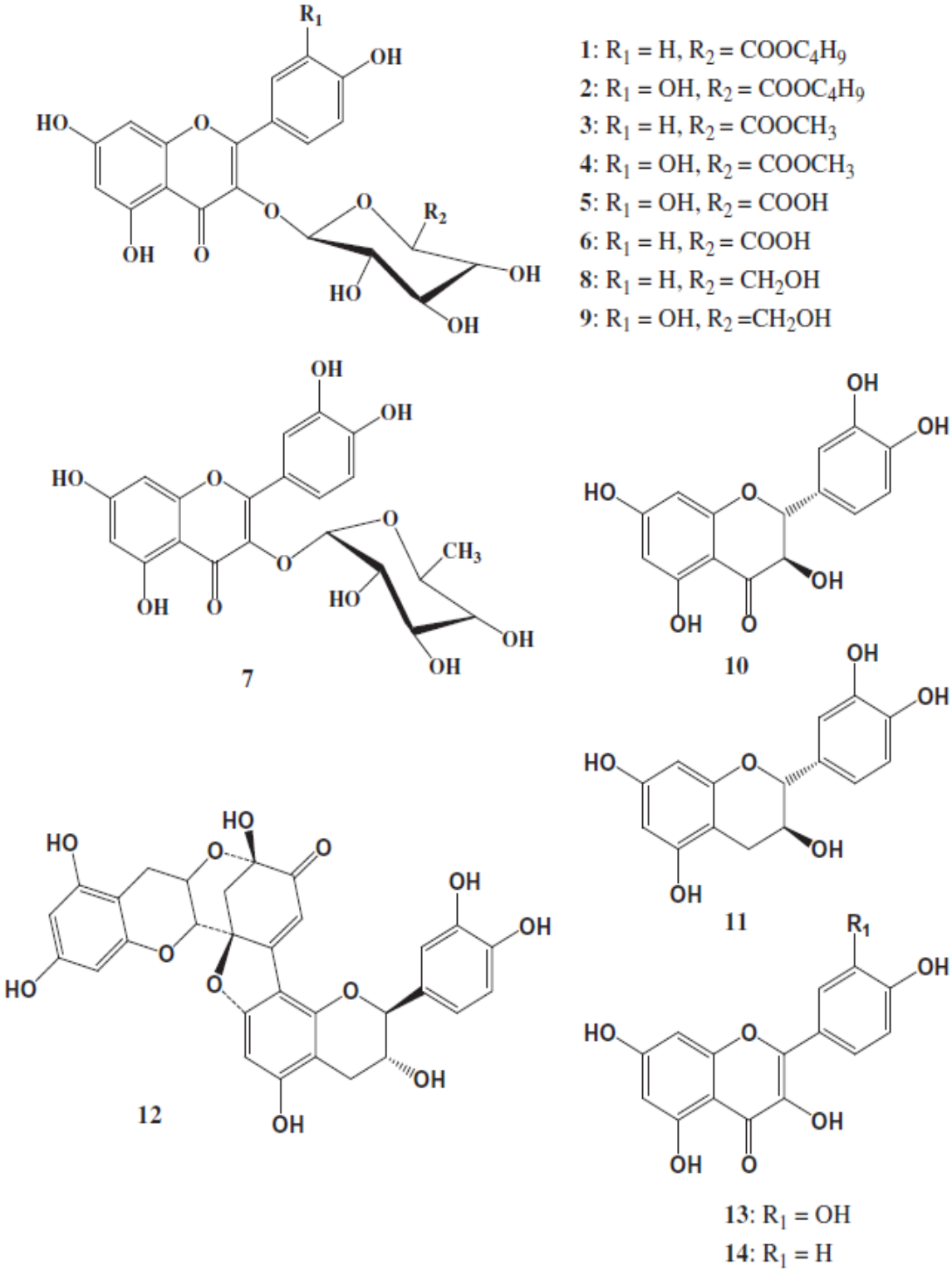


Figure 5.1: The molecular form of each compound.

5.3 Methodology

A dataset of the extracted flavonoids as well as some of the current cancer treatment compounds are shown in Table 5.1.

	Compounds	IC_{50} **Value**
1	Kaempferol-3-*O*-β-D-(6"-n-butyl glucuronide)	17.44
2	Kaempferol-3-*O*-β-D-(6"-methyl glucuronide)	125.54
3	Quercetin 3-*O*-β-D-(6"-n-butyl glucuronide)	60.46
4	Quercetin 3-*O*-β-D-(6"-n-methyl glucuronide)	42.31
5	Quercetin-3-*O*-glucuronide	35.73
6	Kaempferol-3-*O*-glucuronide	64.87
7	Quercetin-3-*O*-rhamnopyranoside	84.17
8	Kaempferol-3-*O*-glucopyranoside	9.21
9	Quercetin-3-*O*-glucopyranoside	66.09
10	Taxifolin	21.25
11	Catechin	20.83
12	Dehydrodicatechin A	131.20
13	Quercetin	4.88
14	Kaempferol	46.18

Table 5.1: Table of compounds and their IC_{50} values.

Each compound is transformed into its molecular graph form to be able to calculate the leaf number and average distance. Figure 5.2 shows the molecular graph for the compound Quercetin (Compound 13 in Table 5.1). The full list of molecular graphs for each compound is found in Appendix A. A vertex u has been chosen in Figure 5.2 to allow for the illustration of how the Wiener index is calculated for Quercetin.

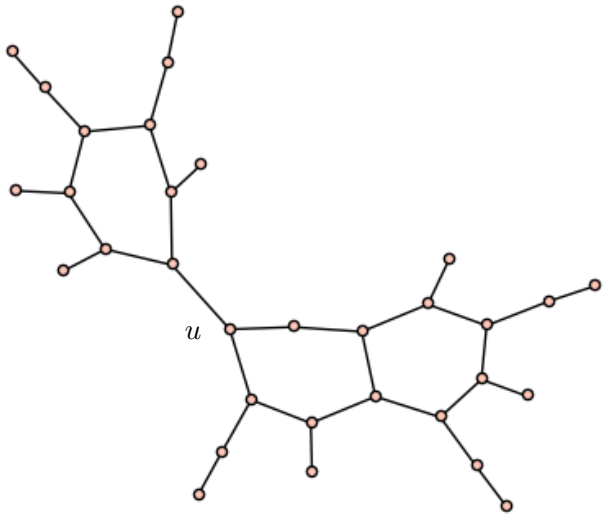


Figure 5.2: The molecular graph of Quercetin.

Letting the molecular graph of Quercetin be G. The Wiener index is calculated by considering $G = G_1 \underset{u_1=u_2}{\cup} G_2$ having been identified at u. Theorem 4.1 is applied to give

$$W(G) = W(G_1) + W(G_2) + (14-1)D_{G_2}(u_2) + (19-1)D_{G_1}(u_1).$$

Graph G_1 has a Wiener index of $W(G_1) = 271$, G_2 has a Wiener index of $W(G_2) = 618$, u_1 has $D_{G_1}(u_1) = 45$, and u_2 has $D_{G_2}(u_2) = 65$. Thus, G has a Wiener index of

$$\begin{aligned} W(G) &= 271 + 618 + (14-1)65 + (19-1)45 \\ &= 2544. \end{aligned}$$

To obtain the average distance, divide $W(G) = 2544$ by $\binom{32}{2}$ which results in

$$\begin{aligned} \mu(G) &= 2544 \times \frac{1}{\binom{32}{2}} \\ &= 5.1290. \end{aligned}$$

Similarly, the average distance is calculated for every compound's molecular graph and

these values are tabulated in Column 2 of Table 5.2. The leaf number is harder to calculate. It requires constructing all possible spanning trees of each molecule followed by checking which one has the largest number of leaves. The leaf number for each molecular graph is calculated using the SageMaths software package and the results are tabulated in column 3 of Table 5.2.

Compound	**Leaf Number**	**Average Distance**
1	30	7.4669
2	30	7.5320
3	26	7.0494
4	26	7.1153
5	22	6.6101
6	22	6.5396
7	24	6.5008
8	24	6.5000
9	24	6.5718
10	15	5.0642
11	16	5.0487
12	33	7.1010
13	13	5.1290
14	13	5.0280

Table 5.2: Table of compounds with calculated leaf number and average distance.

5.4 Results

Shown in Table 5.3 is the relationship between the IC_{50} value and the leaf number. Table 5.4 shows the relationship between the IC_{50} value and the average distance as well as the relationship between the IC_{50} value and the transformed average distance. The transformed average distance is the average distance multiplied by the constant 0.5, all raised to the natural number e. This choice of transformation of the data aligns with the methodology of the papers [24] and [8].

Compound	Predicted IC_{50} Value	Actual IC_{50} Value	Leaf Number
1	80.3520	17.44	30
2	80.3520	125.54	30
3	64.8712	60.46	26
4	64.8712	42.31	26
5	49.3904	35.73	22
6	49.3904	64.87	22
7	57.1308	84.17	24
8	57.1308	9.21	24
9	57.1308	66.09	24
10	22.2990	21.25	15
11	26.1692	20.83	16
12	91.9626	131.20	33
13	14.5586	4.88	13
14	14.5586	46.18	13

Table 5.3: Relationship of IC_{50} value and leaf number.

Shown in Table 5.4 is the relationship between leaf number and IC_{50} value.

Compound	Predicted IC_{50} Value	Actual IC_{50} Value	Average Distance	Transformed Average Distance
1	75.9871	17.44	8.3517	88.2470
2	78.2398	125.54	8.4071	91.7622
3	67.1979	60.46	7.9591	68.6925
4	69.4581	42.31	8.0508	71.4630
5	58.3546	35.73	7.6003	52.7762
6	56.0082	64.87	7.5051	50.5903
7	55.6607	84.17	7.4910	49.4262
8	54.7487	9.21	7.4540	49.4024
9	57.1346	66.09	7.5508	51.5772
10	20.5387	21.25	6.0660	20.8743
11	18.2071	20.83	5.9714	20.6811
12	67.6367	131.2	7.9769	70.8525
13	26.9124	4.88	6.3246	21.7019
14	24.1593	46.18	6.2129	20.4258

Table 5.4: Relationship of IC_{50} value, average distance and transformed average distance.

Two regression models were used to fit the IC_{50} data, one using only the leaf number as a parameter and one using the transformed average distance. Calculating Pearson's correlation coefficient between the leaf number and the transformed average distance, a

value of 95.10% is obtained. This is evidence that the leaf number and the transformed average distance are highly correlated with one another and a model that incorporates both is redundant. Table 5.5 displays the linear equation for each regression model along with its Pearson's correlation coefficient.

Graph Parameter	Model Equation	Correlation (%)
Transformed Average Distance	$7.73050 + 0.8543e^{0.5x}$	27.45
Leaf Number	$-35.7540 + 3.8702x$	38.54

Table 5.5: Models for predicting the IC_{50} value.

Figure 5.3 displays the scatterplot of the IC_{50} values against the transformed average distances. Figure 5.4 displays the scatterplot of the IC_{50} values against the leaf number.

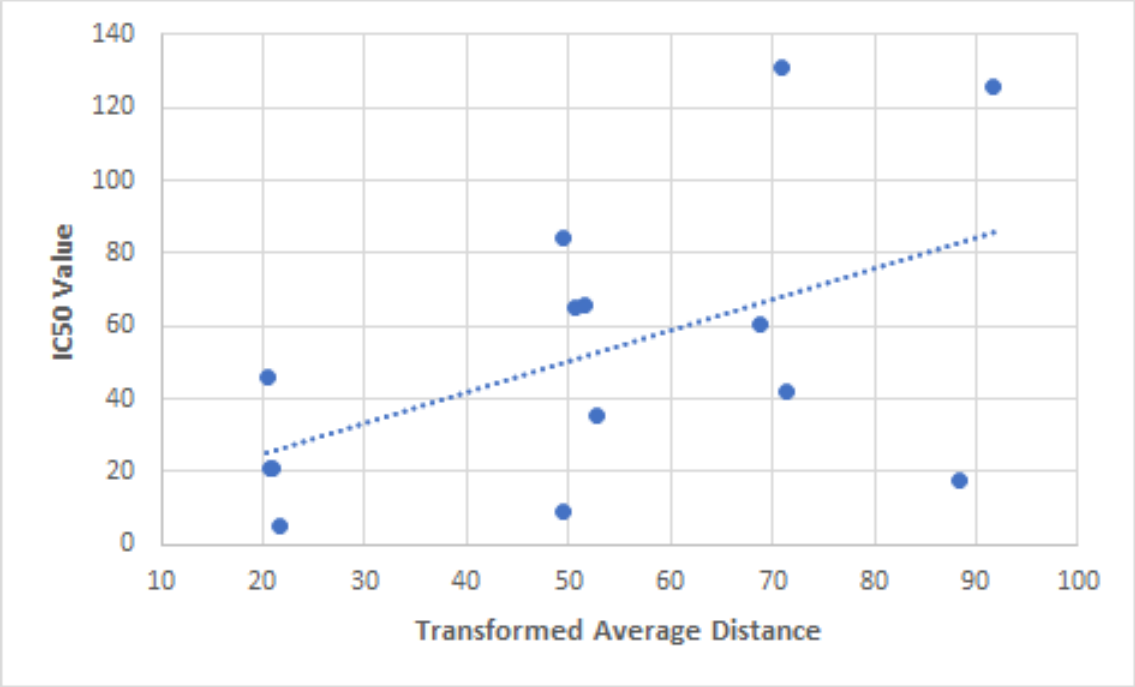


Figure 5.3: Scatterplot of IC_{50} value against the transformed average distance.

In Table 5.3, many compounds have the same leaf number but significantly different IC_{50} values. This shows that leaf number is insufficient, as a single parameter, for modelling IC_{50} value. This is reflected in the poor correlation of 38.54%.

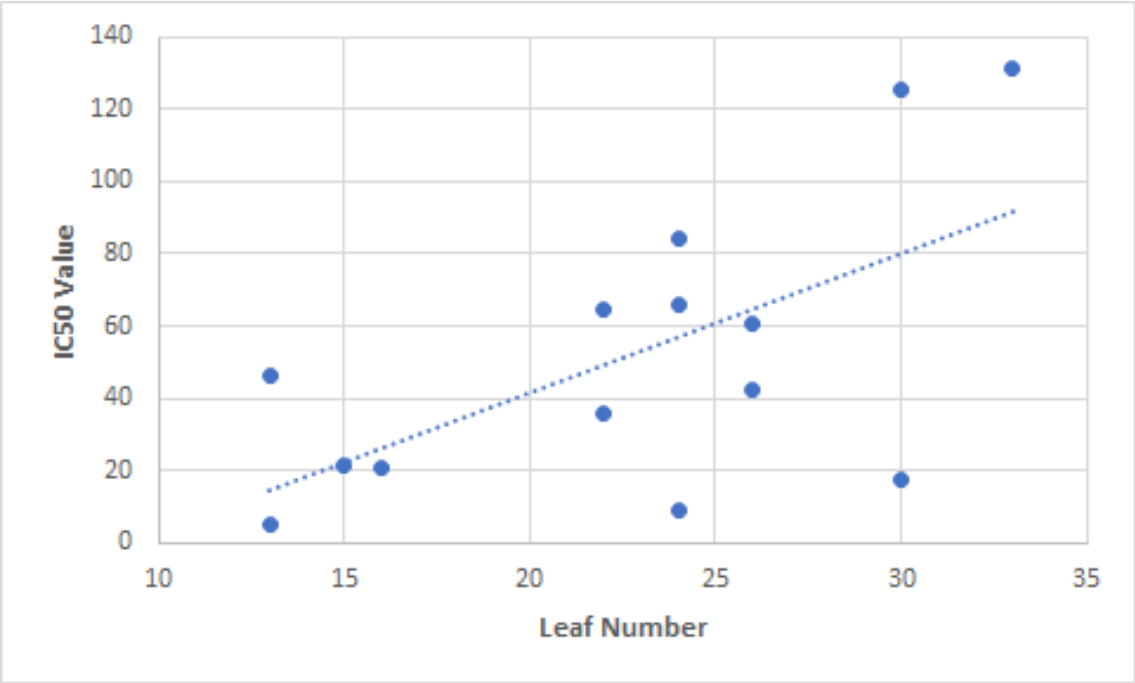


Figure 5.4: Scatterplot of IC_{50} value against the leaf number.

From Table 5.4 we see that compounds one and two have very different IC_{50} values but almost identical transformed average distance. This already tells us that the transformed average distance is a poor predictor, which is reflected by the low correlation of 27.45%. Figure 5.3 displays the scatterplot of the IC_{50} values against the transformed average distances and Figure 5.4 displays the scatterplot of the IC_{50} values against the leaf number. Included in both Figure 5.3 and Figure 5.4 is the line of best fit given by the linear models in Table 5.5.

There are groupings present across both scatterplots. These groupings could be due to the presence of certain structures in the molecule, such as types of elements, that the average distance does not capture. A point raised in [1] is that it appears that compounds containing neighbouring hydroxyls have a lower IC_{50} value. The compounds with this structure present are 2, 4, 5, 7, 9, 10, 11, 12 and 13. Another model, this time using the transformed average distances for these compounds containing the neighbouring hydroxyls, is created and named the *refined model*. The results of this refined model are shown in

Table 5.6.

Refined Model Equation	Correlation (%)
$-24.1950 + 1.5522e^{0.5x}$	59.64

Table 5.6: Refined transformed average distance model.

The refined model has a correlation 59.64%, larger than the 27.45% correlation of the transformed average distance model. Thus, the refined model predicts the data better than the transformed average distance model. There are other structures identified in [1] to have an affect on IC_{50} value, however, there is no evidence of correlation when comparing the transformed average distance for compounds containing these structures.

5.5 Conclusion

Due to the poor correlations for the leaf number and the average distance we conclude that these graph parameters are insufficient as parameters in models that predict the IC_{50} values present in the dataset. Together with the scatterplot of Figure 5.3 and the model of Table 5.6, we conclude that the average distance could be a viable predictor for the dataset in conjunction with other parameters that can incorporate the structural elements found in [1] as well as the groupings present in the scatterplot.

Another point is that our models did not incorporate the double bonds present in any way. The average distance can incorporate these type of bonds through the use of a weighted graph having an edge weight of two for double bonds and 1 for single bonds. We did this but did not show the results in Section 5.4 as there was no significant difference to either the transformed average distance nor the refined average distance.

We conclude that the null hypothesis is correct and reject the alternative hypothesis. Following is the closing chapter for the dissertation.

Chapter 6

Conclusion

In this chapter, we conclude with a summary of the previous findings in this dissertation.

Chapter 1 presents some fundamental Graph Theory concepts, through the use of definitions and examples, needed to understand the proceeding chapters. Some basic Chemistry concepts are also introduced.

Chapter 2 introduces the literature that presents the background to the problems investigated for both the leaf number and the Wiener index.

Chapter 3 puts forward the upper bound and the lower bound on the leaf number. The lower bound on the leaf number, found in Theorem 3.1, is shown to be

$$L(G) \geq \frac{n}{4} + \frac{3}{2},$$

for the specific case of $\delta = 3$ and,

$$L(G) \geq \frac{\delta - 2}{2\delta - 1} n + \frac{2(\delta + 1)}{2\delta - 1}$$

for general δ, found in Theorem 3.2. A counterexample to the work of [11], concerning the upper bound on the leaf number, was presented. A correct upper bound on the leaf number,

$$L(G) \leq n - 1,$$

was given, in Theorem 3.4, and shown to be sharp for any minimum degree δ.

Chapter 4 examines the lower bound and the upper bound for unicyclic graphs and tricyclic graphs,respectively. It is shown, in Theorem 4.1, that for a graph $G = G_1 \underset{u_1=u_2}{\cup} G_2$,

$$W(G) = W(G_1) + W(G_2) + (|V(G_1)| - 1)D_{G_2}(u_2) + (|V(G_2)| - 1)D_{G_1}(u_1).$$

This result is used throughout the chapter to prove the various Wiener index bounds, some of which are our original proofs. The graph $G^0(n,t)$ is defined and shown to be the graph having the smallest Wiener index across all t-cyclic graphs. It is shown, in Theorem 4.2, that for a t-unicyclic graph G, where $|V(G)| = n \geq 4$,

$$W(G) \geq (n-1)^2 - t.$$

Equality is met if and only if $G \cong G^0(n,t)$. The extremal graph using graph transformations is specified. A combinatorial approach is used to identify $G^0(n,3)$. Theorem 4.3 proves that a unicyclic graph G of order $n \geq 4$ has

$$W(G) \leq \frac{1}{6}(n^3 - 7n + 12)$$

with equality for $G \cong S_{n,0,1}$. Theorem 4.5 shows that for G a tricyclic graph of order $n \geq 5$,

$$W(G) \leq \frac{1}{6}(n^3 - 19n + 48)$$

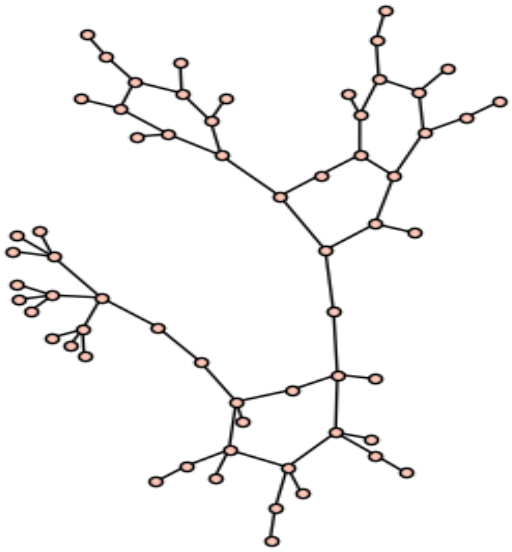
with equality for $G \cong J(n)$.

The lower bound and upper bound on the Wiener index for both unicyclic graphs and tricyclic graphs, respectively, is found and then transformed into bounds for the average distance. We also put forward a t-cyclic set of graphs having the largest Wiener index for any given order.

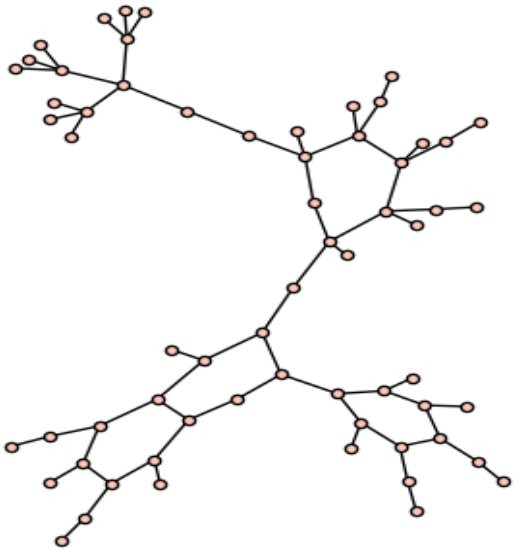
Chapter 5 sets up three models to predict the IC_{50} values of compounds presented in [1]. We conclude that the leaf number is a poor indicator due to the leaf number not being unique across the dataset. The model using the transformed average distance had even worse predicted capabilities. However, refining this model to those compounds that contain neighbouring hydroxyls showed sufficient predictive capabilities such that further research is warranted.

Appendix A

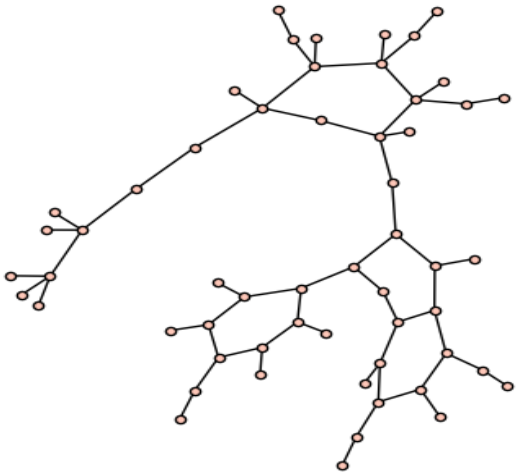
Presented here is the molecular graphs of compounds in Table 5.1 by [1].



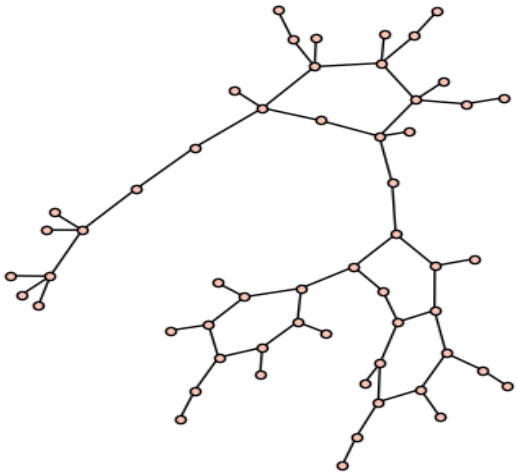
Compound 1.



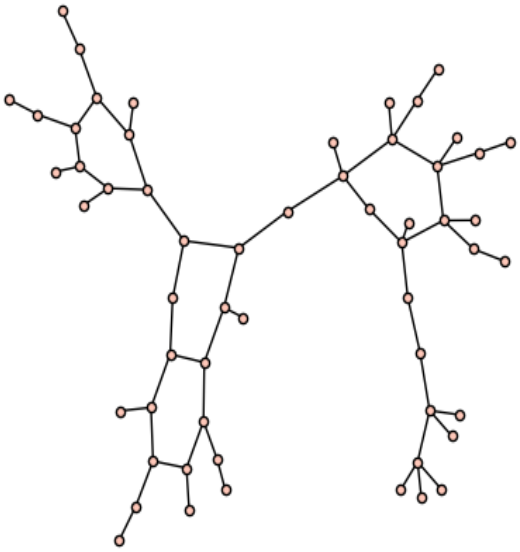
Compound 2.



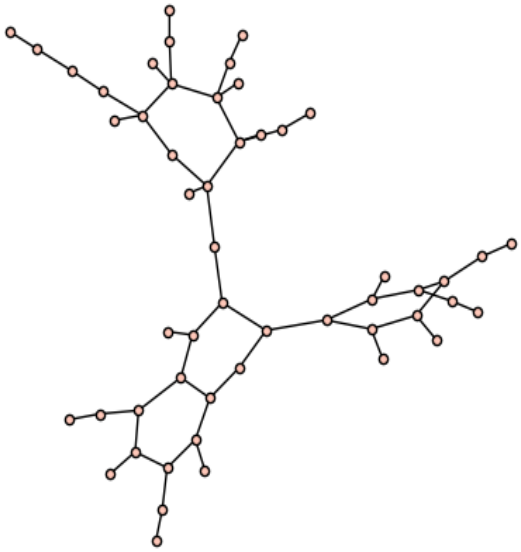
Compound 3.



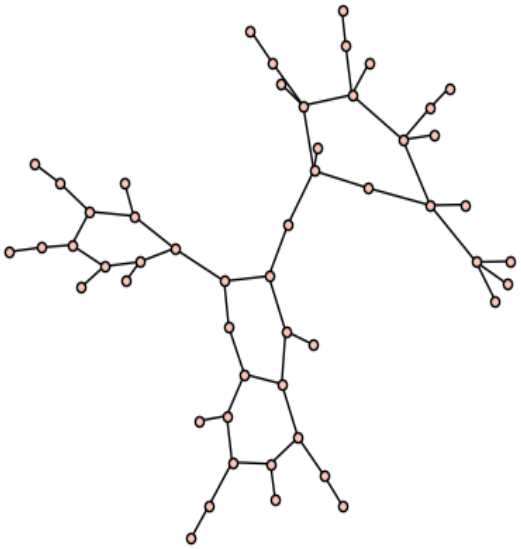
Compound 4.



Compound 5.

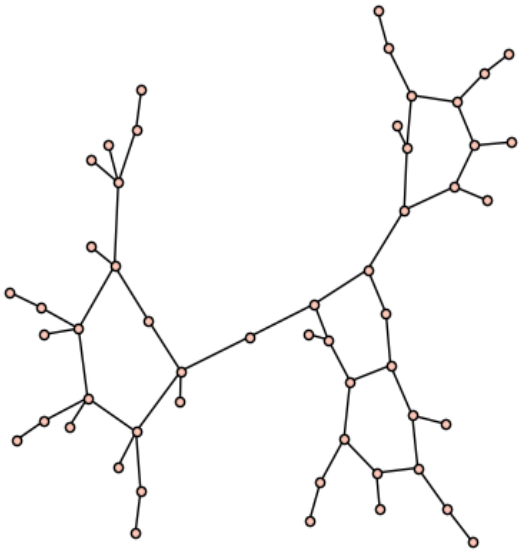


Compound 6.

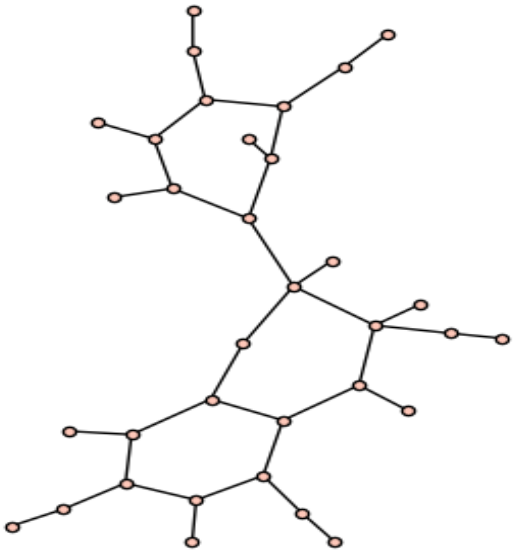


Compound 7.

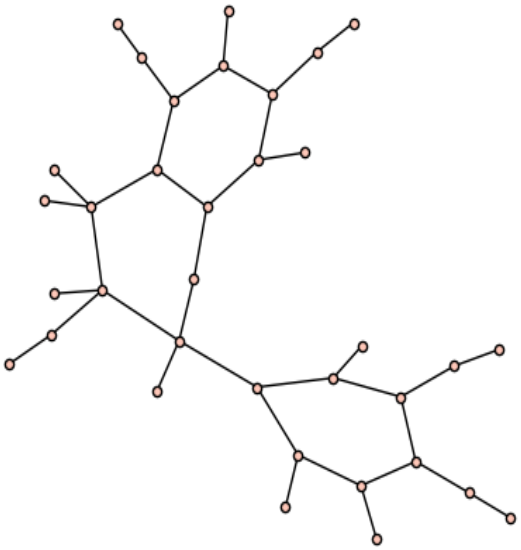
Compound 8.



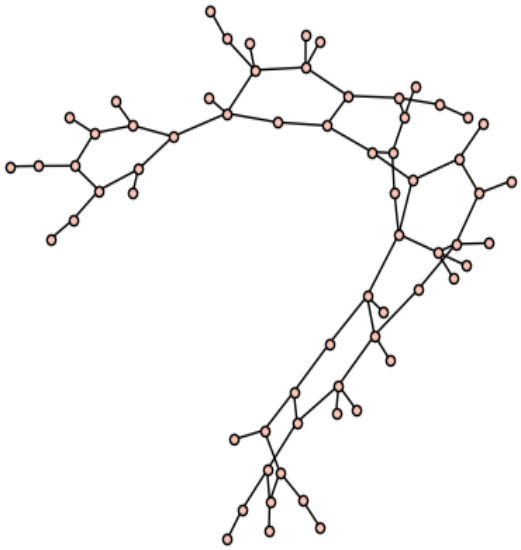
Compound 9.



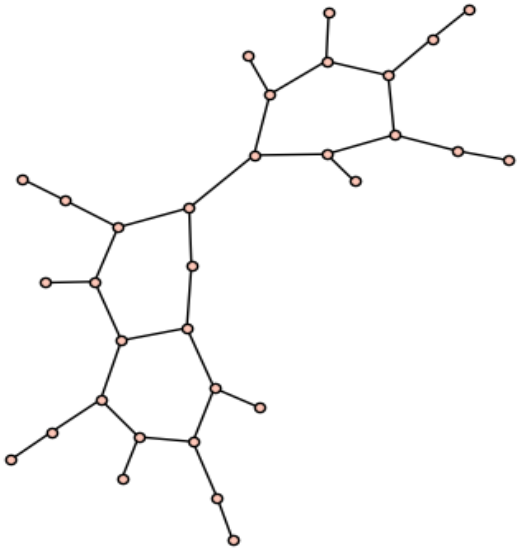
Compound 10.



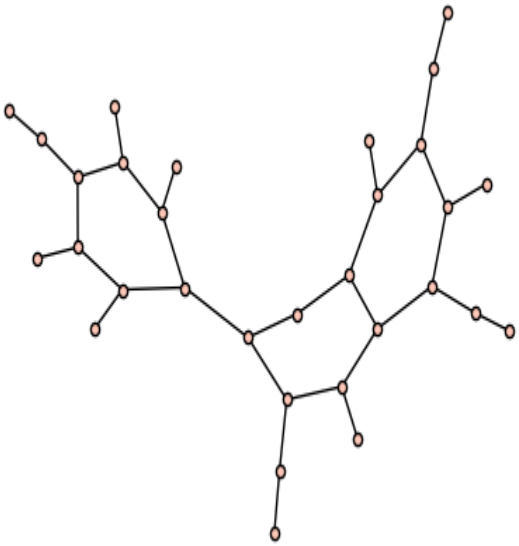
Compound 11.



Compound 12.



Compound 13.



Compound 14.

Bibliography

[1] Ahmed, H., Moawad, A., Owis, A., AbouZid, S., and Ahmed, O. (2016). Flavonoids of calligonum polygonoides and their cytotoxicity. *Pharmaceutical Biology*, 54(10):2119–2126.

[2] Alon, N. (1990). Transversal numbers of uniform hypergraphs. *Graphs and Combinatorics*, 6(1):1–4.

[3] Balaban, A. T. (1985). Applications of graph theory in chemistry. *Journal of chemical information and computer sciences*, 25(3):334–343.

[4] Caro, Y., West, D. B., and Yuster, R. (2000). Connected domination and spanning trees with many leaves. *SIAM Journal on Discrete Mathematics*, 13(2):202–211.

[5] Chartrand, G., Lesniak, L., and Zhang, P. (1996). *Graphs & digraphs*, volume 22. Chapman & Hall London.

[6] Dankelmann, P. and Entringer, R. (2000). Average distance, minimum degree, and spanning trees. *Journal of Graph Theory*, 33(1):1–13.

[7] Griggs, J. R. and Wu, M. (1992). Spanning trees in graphs of minimum degree 4 or 5. *Discrete Mathematics*, 104(2):167–183.

[8] Gupta, S., Singh, M., and Madan, A. (2002). Eccentric distance sum: A novel graph invariant for predicting biological and physical properties. *Journal of Mathematical Analysis and Applications*, 275(1):386–401.

[9] Gutman, I., Li, S., and Wei, W. (2017). Cacti with n-vertices and t cycles having extremal wiener index. *Discrete Applied Mathematics*, 232:189–200.

[10] Haynes, T. W., Hedetniemi, S. T., and Henning, M. A. (2020). *Topics in domination in graphs*, volume 64. Springer.

[11] Kleitman, D. J. and West, D. B. (1991). Spanning trees with many leaves. *SIAM Journal on discrete mathematics*, 4(1):99–106.

[12] Kozlowska, A. and Szostak-Wegierek, D. (2014). Flavonoids-food sources and health benefits. *Roczniki Państwowego Zakładu Higieny*, 65(2).

[13] Kumar, V. and Madan, A. (2007). Application of graph theory: Models for prediction of carbonic anhydrase inhibitory activity of sulfonamides. *Journal of mathematical chemistry*, 42(4):925–940.

[14] Liu, H. and Lu, M. (2007). A unified approach to extremal cacti for different indices. *MATCH Commun. Math. Comput. Chem*, 58(1):183–194.

[15] Mafuta, P., Mukwembi, S., and Munyira, S. (2023). Radius, leaf number, connected domination number and minimum degree. *Quaestiones Mathematicae*, 46(5):1009–1016.

[16] Mukwembi, S. and Nyabadza, F. (2021). A new model for predicting boiling points of alkanes. *Scientific reports*, 11(1):24261.

[17] Mukwembi, S. and Nyabadza, F. (2023). Predicting anti-cancer activity in flavonoids: a graph theoretic approach. *Scientific Reports*, 13(1):3309.

[18] Payan, C., Tchuente, M., and Xuong, N. H. (1984). Arbres avec un nombre maximum de sommets pendants. *Discrete Mathematics*, 49(3):267–273.

[19] Polansky, O. and Bonchev, D. (1986). The wiener number of graphs. i. general theory and changes due to some graph operations. *MATCH Communications in Mathathematics and in Computer Chemistry*, 21(133-186):72.

[20] Rauf, A., Imran, M., Khan, I. A., ur Rehman, M., Gilani, S. A., Mehmood, Z., and Mubarak, M. S. (2018). Anticancer potential of quercetin: A comprehensive review. *Phytotherapy research*, 32(11):2109–2130.

[21] Rouvray, D. H. (1986). Predicting chemistry from topology. *Scientific American*, 255(3):40–47.

[22] Rouvray, D. H. (2002). The rich legacy of half a century of the wiener index. In *Topology in Chemistry*, pages 16–37. Elsevier.

[23] Sampathkumar, E. and Walikar, H. (2016). The connected domination number of a graph. *Journal of Mathematical and Physical Sciences*, 13.

[24] Sharma, V., Goswami, R., and Madan, A. (1997). Eccentric connectivity index: A novel highly discriminating topological descriptor for structure- property and structure-activity studies. *Journal of chemical information and computer sciences*, 37(2):273–282.

[25] Šoltés, L. (1991). Transmission in graphs: a bound and vertex removing. *Mathematica Slovaca*, 41(1):11–16.

[26] Storer, J. A. (1981). Constructing full spanning trees for cubic graphs. *Information Processing Letters*, 13(1):8–11.

[27] Wang, D., Tan, S., and Zhu, L. (2016). On the lower and upper bounds for different indices of tricyclic graphs. *Journal of Applied Mathematics and Computing*, 51(1-2):1–11.

[28] Wiener, H. (1947). Structural determination of paraffin boiling points. *Journal of the American chemical society*, 69(1):17–20.

[29] Xhanti, S. (2021). On the wiener index of bicyclic graphs and graphs with fixed segment sequence.

[30] Yu, G. and Feng, L. (2010). On the wiener index of unicyclic graphs with given girth. *Ars Combinatoria*, 94:361–369.